(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
7 August 2003 (07.08.2003)

PCT

(10) International Publication Number
WO 03/064591 A2

(51) International Patent Classification[7]: C12N

(21) International Application Number: PCT/US02/33496

(22) International Filing Date: 18 October 2002 (18.10.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/351,004 19 October 2001 (19.10.2001) US

(71) Applicants *(for all designated States except US)*: **THE GOVERNMENT OF THE UNITED STATES OF AMERICA as represented by THE SECRETARY OF HEALTH AND HUMAN SERVICES** [US/US]; The National Institutes of Health, Office of Technology Transfer, 6011 Executive Boulevard, Suite 325, Rockville, MD 20852-3804 (US). **AVENTIS PHARMA** [FR/FR]; 112, rue de Bry, F-94460 Chennevieres sur Marne (FR).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **DEAN, Michael, C.** [US/US]; 1362 Hitchingpost Lane, Frederick, MD 21703 (US). **SCHRIML, Lynn** [US/US]; 204 Thames Drive, Frederick, MD 21702 (US). **ALLIKMETS, Rando, L.** [ES/US]; 200 Maple Road, Cornwall-on-Hudson, NY 12520 (US). **DENEFLE, M., Patrice** [FR/FR]; 45, avenue des FUsilles de Château, F-94100 Saint Maur (FR). **ROSIER-MONTUS, Marie-Francoise** [FR/FR]; 21, rue des Baconnets, F-92160 Antony (FR). **ARNOULD-REGUIGNE, Isabelle** [FR/FR]; 112 rue de Brie, F-94430 Chennevieres sur Marne (FR). **PRADES, CAtherine** [FR/FR]; 30 Avenue du Général de Gaulle, F-94320 Thiais (FR).

(74) Agent: **HARDING, Tanya, M.**; Klarquist Sparkman, LLP, One World Trade Center, Suite 1600, 121 SW Salmon Street, Portland, OR 97204 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/064591 A2

(54) Title: ABCA8 NUCLEIC ACIDS AND PROTEINS, AND USES THEREOF

(57) Abstract: The present invention relates to nucleic acids corresponding to the ABCA8 gene and cDNA, the ABCA8 protein, and methods of making and using these molecules. Also provided are methods of ameliorating, treating, detecting, prognosing, and diagnosing diseases and conditions believed to be associated with altered expression o fABCA8, such as hypercholesterolemia, drug resistance, retinal degeneration, or neurological disease. Kits and pharmaceutical compositions are also provided.

ABCA8 NUCLEIC ACIDS AND PROTEINS, AND USES THEREOF

PRIORITY CLAIM

This application claims the benefit of U.S. Provisional Application No. 60/351,004, filed October 19, 2001, which is incorporated by reference in its entirety herein.

FIELD

The present disclosure is related to extra- and intra-cellular transport, including the mechanisms controlling transport, diseases that arise from defects in such mechanisms, and methods of influencing (either inhibiting, enhancing or otherwise changing) transport.

BACKGROUND

The ABC (ATP-binding cassette transporter) gene superfamily encodes active transporter proteins and constitutes a family of proteins that have been extremely well conserved through evolution, from bacteria to humans (Ames and Lecar, *FASEB J.*, 1992, 6, 2660-2666). The ABC proteins are involved in extra- and intracellular membrane transport of various substrates, for example ions, amino acids, peptides, sugars, vitamins, or steroid hormones. Among the 40 characterized human ABC genes members, 11 members have been described as associated with human disease, such as ABCA1, ABCA4 (ABCR) and ABCC7, (CFTR) which are thought to be involved in Tangier disease (Bodzioch *et al.*, *Nat. Genet.*, 1999, 22(4):347-351; Brooks-Wilson *et al.*, *Nat Genet*, 1999, 22(4), 336-345 ; Rust *et al.*, *Nat. Genet.*, 1999, 22, 352-355), the Stargardt disease (Lewis *et al.*, *Am. J. Hum. Genet.*, 1999, 64, 422-434), and Cystic Fibrosis (Riordan *et al.*, *Science*, 1989, 245, 1066-1073), respectively. Another ABC transporter, *mdr1* (multidrug resistance) gene, encodes the protein *mdr1a*, also called P-glycoprotein (P-gp). This protein functions as a drug-efflux transmembrane protein pump. P-glycoprotein was first identified over 20 years ago in chemotherapeutic drug-resistant tumor cells, and is now known to be a major cause of multidrug resistance in many cancers (Van Asperen *et al.*, *J. Pharmaceut. Sci.* 86:881-884, 1997, 1997; Tsuji, *Therap. Drug Monitor.* 20:588-590, 1998). These implications reveal the importance of the functional role of the ABC gene family. The discovery of a new family of ABC gene members is expected to provide new insights into the physiopathology and treatment of human diseases.

The prototype ABC protein binds ATP and uses the energy from ATP hydrolysis to drive the transport of various molecules across cell membranes. The functional protein contains two ATP-binding domains (nucleotide binding fold, NBF) and two transmembrane (TM) domains. The genes are typically organized as full transporters containing two of each domain, or half transporters with only one of each domain. Most full transporters are arranged in a TM-NBF-TM-NBF fashion (Dean *et al.*, *Curr Opin Genet*, 1995, 5, 79-785).

Analysis of amino acids sequence alignments of the ATP-binding domains has allowed the ABC genes to be separated into sub-families (Allikmets *et al.*, *Hum Mol Genet*, 1996, 5, 1649-1655).

Currently, according to the recent HUGO classification, seven ABC gene sub-families named ABC (A to G) have been described in the human genome (ABC1, CFTR/MRP, MDR, ABC8, ALD, GCN20, OABP) with all except one (OABP) containing multiple members. Among the ABC sub-families, the ABCA gene subfamily is probably the most evolutionary complex. The *ABCA* genes and *OABP* represent the only two sub-families of ABC genes that do not have identifiable orthologs in the yeast genome (Decottignies and Goffeau, *Nat Genet.* Feb;15(2):137-45,1997; Michaelis and Berkower, *Cold Spring Harb Symp Quant Biol.* 60:291-307, 1995). There is, however, at least one *ABCA*-related gene in *C. elegans* (*ced-7*) and several in *Drosophila*. Thus, the ABCA genes appear to have diverged after eukaryotes became multicellular and developed more sophisticated transport requirements.

ABCA1 was demonstrated to be the gene responsible for Tangier disease, a disorder characterized by high levels of cholesterol in peripheral tissues, and a very low level of HDLs, and familial hypoalphalipoproteinemia (FHD) (Bodzioch *et al.*, *Nat Genet* (1999) 22, 347-51; Brooks-Wilson *et al.*, *Nat Genet* (1999) 336-45; Rust *et al.*, *Nat Genet* (1999) 22, 352-5; Marcil *et al.*, *The Lancet* (1999) 354, 1341-46). The ABCA1 protein is proposed to function in the reverse transport of cholesterol from peripheral tissues via an interaction with the apolipoprotein 1 (ApoA-1) of HDL tissues (see Wang *et al.*, *JBC* 275(42): 33053-33058, 2000).

The *ABCA2* gene is highly expressed in the brain, and *ABCA3* in the lung, but no function has been ascribed to these loci. The *ABCA4* gene is exclusively expressed in the rod photoreceptors of the retina and mutations thereof are responsible for several pathologies of human eyes, such as retinal degenerative disorders (Allikmets *et al.*, *Science* (1997) 277, 1805-1807; Allikmets *et al.*, *Nat Genet* (1997) 15, 236-246; Sun *et al.*, *J Biol Chem* (1999) 8269-81; Weng *et al.*, *Cell* (1999) 98, 13-23; Cremers *et al.*, *Hum Mol Genet* (1998) 7, 355-362; Martinez-Mir *et al.*, *Genomics* (1997) 40, 142-146). *ABCA4* is believed to transport retinal and/or retinal-phospholipid complexes from the rod photoreceptor outer segment disks to the cytoplasm, facilitating phototransduction.

Characterization of new genes from the ABCA subfamily is likely to yield biologically important transporters, which may have translocase activity for membrane lipid transport or the transport of other substances and which may play a role in human pathologies.

SUMMARY OF THE DISCLOSURE

This disclosure describes the discovery and characterization of a new gene belonging to the ABCA protein sub-family, which has been designated *ABCA8*. This new gene appears to be closely related to other ABCA subfamily members such as ABCA1. ABCA8 is proposed to be involved in the energy-dependent transport of one or a variety of substances, for example ions, metals, amino acids, lipids, peptides, sugars, vitamins and steroid hormones. The newly discovered gene also shows considerable conservation of the amino acid sequences when compared to previously known ABCA genes, particularly within the transmembrane region (TM) and the ATP-binding regions (NBF), and has a similar gene organization to *ABCA1*.

ABCA8 is one of a group of several ABCA genes that is organized in a single large cluster on chromosome 17q24, in a head-to-tail fashion. Furthermore, *ABCA8* is transcribed with a tissue-specific distribution, but presents a heterogeneous pattern of expression, suggesting a regional and possibly functional specialization of the corresponding proteins.

The present disclosure relates to novel nucleic acids encoding ABCA8, for instance the *ABCA8* gene (SEQ ID NO: 1) and the predicted amino acid sequence (SEQ ID NO: 2). The protein is believed to be involved in the reverse transport of cholesterol, as well as in the membrane transport of other lipophilic molecules, for instance inflammation mediating substances such as prostaglandins and prostacyclins, and in any pathology whose candidate chromosomal region is situated on chromosome 17, more precisely on the 17q arm and still more precisely in the 17q24 locus. The disclosure also relates to means for the detection of polymorphisms in general, and mutations in particular, in the *ABCA8* gene or corresponding protein produced by allelic forms of the *ABCA8* gene.

Also disclosed are methods of using these molecules in detecting biological conditions associated with mutation, altered expression, or duplication of *ABCA8* in a subject and methods of screening for agents that modulate ABCA8 transporter activity, such as specific transport inhibitors. Oligonucleotides for use in examples of such methods are also provided.

Also disclosed herein are protein specific binding agents, such as antibodies, that bind specifically to at least one epitope of an ABCA8 variant protein preferentially compared to wildtype ABCA8, and methods of using such antibodies in diagnosis and screening.

Kits are also provided for carrying out the methods described herein.

The foregoing and other features and advantages of the disclosure will become more apparent from the following detailed description of several embodiments, which proceeds with reference to the accompanying Figures.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 represents a map of the 17q24 region containing the *ABCA5, 6, 8-10* genes. A physical map of the portion of chromosome 17q24 is shown in **Figure 1A** containing the 5 *ABCA* genes. In **Figure 1A**, the location of the boundaries of BAC clones hRPK.235_I_10 and hRPK.293_K_20 (GenBank accession #s AC005495, AC005922) is indicated. Gene orientation is indicated by the arrows, and the size and location of the corresponding transcripts is shown below the map. • indicates the initiation codon; | represents the stop codon; and --- symbolizes the working draft sequences. **Figure 1B** shows a prediction of the intron and exon sequence locations within *ABCA5, 10, 6, 9,* and *8* in genomic DNA. Analysis was performed by combination of several gene-finding programs such as GENSCAN (Burge and Karlin, 1997, *J Mol Biol.*; 268(1):78-94), FGENEH/FEXH (Solovyev and Salamov, 1997, *Ismb*; 5:294-302), and XPOUND (Thomas and Skolnick, 1994, *J Math Appl Med Biol.*;11(1):1-16).

Figure 2 represents the alignment of ABC1-like genes. An alignment of the amino acid sequences of the full-length *ABCA10* (huest698739), *8* (huest990006), *9* (huest640918), and *6*

(huest155051) open reading frames, the partial sequence of *ABCA5* (huest90625) and a gene originally called *ABCA11*, that is now known to be a pseudogene (huest1133530) are shown as aligned.

Figure 3 displays two phylogenetic linkage trees of ABC1-like sequences. Phylogenetic trees were constructed with the alignments of the N- and C- terminal ATP-binding domains' sequence by both Neighbor joining (**Figure 3A**) and maximum parsimony (**Figure 3B**) methods.

Figure 4 shows the alignment of the full length ABCA8 cDNA (**Figure 4A**) and protein (**Figure 4B**) with the putative and incomplete ABCA8 cDNA and protein previously published in GenBank (referred to in the Figure as "Ref."; Accession numbers NM_007168 and NP_009099.1, respectively).

SEQUENCE LISTING

The nucleic and amino acid sequences listed in the accompanying sequence Listings are shown using standard letter abbreviations for nucleotide bases, and three letter code for amino acids, as defined in 37 C.F.R. 1.822. Only one strand of each nucleic acid sequence is shown, but the complementary strand is understood as included by any reference to the displayed strand. In the accompanying sequence listing:

SEQ ID NO: 1 shows the nucleic acid of human *ABCA8* cDNA and the corresponding deduced amino acid sequence.

SEQ ID NO: 2 shows the amino acid sequence of the human ABCA8 protein encoded by *ABCA8* (SEQ ID NO: 1).

SEQ ID NO: 3 and SEQ ID NO: 4 show the nucleic acid sequences of two primers that can be used to amplify SEQ ID NO: 1 using the polymerase chain reaction PCR.

SEQ ID NO: 5 and SEQ ID NO: 6 show the nucleic acid sequences of two primers that can be used to amplify the region of SEQ ID NO: 1 corresponding to the ABCA8 protein coding region.

SEQ ID NO: 7 and SEQ ID NO: 8 show the nucleic acid sequences of two primers that can be used to amplify a particular region within ABCA8.

SEQ ID NO: 9 and SEQ ID NO: 10 show the nucleic acid sequences of two primers that can be used to amplify an ABCA8 encoding sequence.

DETAILED DESCRIPTION

I. Abbreviations

ABC:	ATP-binding cassette transporter
BAC:	bacterial artificial chromosome
bp	base pair(s)
DNA:	deoxyribonucleic acid
gDNA:	genomic DNA
pfu:	plaque forming unit

ELISA:	enzyme-linked immunosorbant assay
PCR:	polymerase chain reaction
TAE:	Tris acetate EDTA

II. Terms

Unless otherwise noted, technical terms are used according to conventional usage. Definitions of common terms in molecular biology may be found in Benjamin Lewin, *Genes V*, published by Oxford University Press, 1994 (ISBN 0-19-854287-9); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0-632-02182-9); and Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 1-56081-569-8).

In accordance with the present disclosure, conventional molecular biology, microbiology, and recombinant DNA techniques within the skill of the art are used. Such techniques are fully explained in the literature (Sambrook *et al., Molecular cloning, a laboratory manual,* 2nd ed., Cold Spring Harbor Laboratory, Cold spring Harbor, New York; Glover, 1985, *DNA Cloning: A practical approach*, volumes I and II oligonucleotide synthesis, MRL Press, LTD., Oxford, U.K.; Hames and Higgins, *Gene Transcription: A practical approach Transcription and Translation*, IRL/Oxford University Press, 1993; Freshney, 1986, *Immobilized Cells And Enzymes*, IRL Press; and Perbal, 1984, *A practical guide to molecular cloning*).

In order to facilitate review of the various embodiments of the invention, the following explanations of specific terms are provided:

Altered expression: Expression of a nucleic acid (*e.g.*, mRNA or protein) in a subject or biological sample from a subject that deviates from expression in a subject or biological sample from a subject having normal characteristics for the biological condition associated with the nucleic acid. Normal expression can be found in a control, a standard for a population, etc. For instance, where the altered expression manifests as a transporter disease condition, such as deficient extra- or intracellular transport, characteristics of normal expression might include an individual who is not suffering from the transport disorder, a population standard of individuals believed not to be suffering from the disease, etc. For instance, certain altered expression, such as altered *ABCA8* nucleic acid or ABCA8 protein expression, can be described as being associated with the biological conditions of altered (*e.g.*, reduced) transporter function and tendency to develop a transporter deficiency.

Likewise, altered expression may be associated with a disease. The term "associated with" includes an increased risk of developing the disease as well as the disease itself.

Altered protein expression, such as altered ABCA8 protein expression, also refers to expression of a protein that is in some manner different to expression of the protein in a normal (wild type) situation. This includes but is not necessarily limited to: (1) a mutation in the protein such that one or more of the amino acid residues is different; (2) a short deletion or addition of one or a few amino acid residues to the sequence of the protein; (3) a longer deletion or addition of amino acid

residues, such that an entire protein domain or sub-domain is removed or added; (4) expression of an increased amount of the protein, compared to a control or standard amount; (5) expression of an decreased amount of the protein, compared to a control or standard amount; (6) alteration of the subcellular localization or targeting of the protein; (7) alteration of the temporally regulated expression of the protein (such that the protein is expressed when it normally would not be, or alternatively is not expressed when it normally would be); and (8) alteration of the localized (*e.g.*, organ or tissue specific) expression of the protein (such that the protein is not expressed where it would normally be expressed or is expressed where it normally would not be expressed), each compared to a control or standard.

Animal: Living multi-cellular vertebrate organisms, a category that includes for example, mammals and birds.

Amplification: When used in reference to a nucleic acid, techniques that increase the number of copies of a nucleic acid molecule in a sample or specimen. An example of amplification is the polymerase chain reaction, in which a biological sample collected from a subject is contacted with a pair of oligonucleotide primers, under conditions that allow for the hybridization of the primers to nucleic acid template in the sample. The primers are extended under suitable conditions, dissociated from the template, and then re-annealed, extended, and dissociated to amplify the number of copies of the nucleic acid. The product of *in vitro* amplification can be characterized by electrophoresis, restriction endonuclease cleavage patterns, oligonucleotide hybridization or ligation, and/or nucleic acid sequencing, using standard techniques. Other examples of *in vitro* amplification techniques include strand displacement amplification (*see* U.S. Patent No. 5,744,311); transcription-free isothermal amplification (*see* U.S. Patent No. 6,033,881); repair chain reaction amplification (*see* WO 90/01069); ligase chain reaction amplification (*see* EP-A-320 308); gap filling ligase chain reaction amplification (*see* U.S. Patent No. 5,427,930); coupled ligase detection and PCR (*see* U.S. Patent No. 6,027,889); and NASBA™ RNA transcription-free amplification (*see* U.S. Patent No. 6,025,134).

Antisense, Sense, and Antigene: Double-stranded DNA (dsDNA) has two strands, a 5' -> 3' strand, referred to as the plus strand, and a 3' -> 5' strand (the reverse compliment), referred to as the minus strand. Because RNA polymerase adds nucleic acids in a 5' -> 3' direction, the minus strand of the DNA serves as the template for the RNA during transcription. Thus, the RNA formed will have a sequence complementary to the minus strand and identical to the plus strand (except that U is substituted for T).

Antisense molecules are molecules that are specifically hybridizable or specifically complementary to either RNA or the plus strand of DNA. Sense molecules are molecules that are specifically hybridizable or specifically complementary to the minus strand of DNA. Antigene molecules are either antisense or sense molecules directed to a dsDNA target.

Binding or stable binding: An oligonucleotide binds or stably binds to a target nucleic acid if a sufficient amount of the oligonucleotide forms base pairs or is hybridized to its target nucleic acid, to permit detection of that binding. Binding can be detected by either physical or functional

properties of the target:oligonucleotide complex. Binding between a target and an oligonucleotide can be detected by any procedure known to one skilled in the art, including both functional and physical binding assays. Binding can be detected functionally by determining whether binding has an observable effect upon a biosynthetic process such as expression of a gene, DNA replication, transcription, translation and the like.

Physical methods of detecting the binding of complementary strands of DNA or RNA are well known in the art, and include such methods as DNase I or chemical footprinting, gel shift and affinity cleavage assays, Northern blotting, dot blotting and light absorption detection procedures. For example, one widely method involves observing a change in light absorption of a solution containing an oligonucleotide (or an analog) and a target nucleic acid at 220 to 300 nm as the temperature is slowly increased. If the oligonucleotide or analog has bound to its target, there is a sudden increase in absorption at a characteristic temperature as the oligonucleotide (or analog) and the target disassociate from each other, or melt.

The binding between an oligomer and its target nucleic acid is frequently characterized by the temperature (T_m) at which 50% of the oligomer is melted from its target. A higher (T_m) means a stronger or more stable complex relative to a complex with a lower (T_m).

Biological condition: Designates a condition of a subject that can be assessed through observation or through the analysis of a biological sample, *e.g.*, altered expression level of ABCA8 protein in comparison to a control expression level, or ability of cells from a subject to transport cholesterol.

Biological sample: Any sample in which the presence of a protein and/or ongoing expression of a protein may be detected. Suitable biological samples include samples containing genomic DNA or RNA (including mRNA), obtained from body cells of a subject, such as but not limited to those present in peripheral blood, urine, saliva, tissue biopsy, surgical specimen, amniocentesis samples and autopsy material.

Cassette: A segment of DNA that can be inserted into a vector at specific restriction sites. In some embodiments, the segment of DNA encodes a polypeptide of interest, and the cassette and restriction sites are designed to ensure insertion of the cassette in the proper reading frame for transcription and translation.

cDNA (complementary DNA): A piece of DNA lacking internal, non-coding segments (introns) and transcriptional regulatory sequences. cDNA can also contain untranslated regions (UTRs) that are responsible for translational control in the corresponding RNA molecule. cDNA is synthesized in the laboratory by reverse transcription from messenger RNA extracted from cells.

Coding sequence: A DNA sequence that is transcribed and translated into a polypeptide in a cell *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence are determined by a start codon at the 5' (amino) terminus and a translation stop codon at the 3' (carboxyl) terminus. A coding sequence can include, but is not limited to, prokaryotic sequences, cDNA from eukaryotic mRNA, genomic DNA sequences from

eukaryotic (*e.g.*, mammalian) DNA, and even synthetic DNA sequences. If the coding sequence is intended for expression in a eukaryotic cell, a polyadenylation signal and transcription termination sequence will usually be located 3' to the coding sequence.

Transcriptional and translational control sequences include DNA regulatory sequences, such as promoters, enhancers, terminators, and so forth, which provide for the expression of a coding sequence in a host cell. In eukaryotic cells, polyadenylation signals are control sequences. A coding sequence is under the control of transcriptional and translational control sequences in a cell when RNA polymerase transcribes the coding sequence into mRNA, which is then trans-RNA spliced and translated into the protein encoded by the coding sequence.

Corresponding to is used herein to refer to similar or homologous sequences, whether the exact position is identical or different from the molecule to which the similarity or homology is measured. A nucleic acid or amino acid sequence alignment may include spaces. Thus, the term "corresponding to" refers to the sequence similarity, and not the numbering of the amino acid residues or nucleotide bases.

DNA (deoxyribonucleic acid): A long chain polymer that comprises the genetic material of most living organisms (some viruses have genes comprising ribonucleic acid (RNA)). The repeating units in DNA polymers are four different nucleotides, each of which comprises one of the four bases, adenine, guanine, cytosine and thymine bound to a deoxyribose sugar to which a phosphate group is attached. Triplets of nucleotides (referred to as codons) code for each amino acid in a polypeptide. The term codon is also used for the corresponding (and complementary) sequences of three nucleotides in the mRNA into which the DNA sequence is transcribed.

Unless otherwise specified, any reference to a DNA molecule is intended to include the reverse complement of that DNA molecule. Except where single-strandedness is required by the text herein, DNA molecules, though written to depict only a single strand, encompass both strands of a double-stranded DNA molecule. Thus, a reference to the nucleic acid molecule that encodes a specific protein, or a fragment thereof, encompasses both the sense strand and its reverse complement. Thus, for instance, it is appropriate to generate probes or primers from the reverse complement sequence of the disclosed nucleic acid molecules.

Deletion: The removal of a sequence of DNA, the regions on either side being joined together.

Encode: A polynucleotide is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for and/or the polypeptide or a fragment thereof. The anti-sense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

Functional fragments and variants of a polypeptide: Included are those fragments and variants that maintain at least one function of the parent polypeptide. It is recognized that the gene or cDNA encoding a polypeptide can be considerably mutated without materially altering one or more the polypeptide's functions. First, the genetic code is well known to be degenerate, and thus different

codons encode the same amino acids. Second, even where an amino acid substitution is introduced, the mutation can be conservative and have no material impact on the essential functions of a protein (see Stryer, *Biochemistry* 4th Ed., W. Freeman & Co., New York, NY, 1995). Third, part of a polypeptide chain can be deleted without impairing or eliminating all of its functions. *e.g.*, sequence variants a protein, such as a 5' or 3' variant, may retain the full function of an entire protein. Fourth, insertions or additions can be made in the polypeptide chain for example, adding epitope tags, without impairing or eliminating its functions (Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1998). Other modifications that can be made without materially impairing one or more functions of a polypeptide include, for example, *in vivo* or *in vitro* chemical and biochemical modifications or the incorporation of unusual amino acids. Such modifications include, for example, acetylation, carboxylation, phosphorylation, glycosylation, ubiquination, labeling, *e.g.*, with radionucleides, and various enzymatic modifications, as will be readily appreciated by those well skilled in the art. A variety of methods for labeling polypeptides and labels useful for such purposes are well known in the art, and include radioactive isotopes such as ^{32}P, ligands that bind to or are bound by labeled specific binding partners (*e.g.*, antibodies), fluorophores, chemiluminescent agents, enzymes, and antiligands. Functional fragments and variants can be of varying length. For example, a fragment may consist of 10 or more, 25 or more, 50 or more, 75 or more, 100 or more, or 200 or more amino acid residues.

A functional fragment or variant of ABCA8 is defined herein as a polypeptide that is capable of transporter activity, including any polypeptide six or more amino acid residues in length that is capable of transporter activity.

Gene refers to an assembly of nucleotides that encode a polypeptide, and includes cDNA and genomic DNA corresponding to a polypeptide.

Heterologous DNA refers to DNA not naturally located in the cell, or in a chromosomal site of the cell. In some embodiments, the heterologous DNA includes a gene foreign to the cell.

Hybridization: Oligonucleotides and their analogs hybridize by hydrogen bonding, which includes Watson-Crick, Hoogsteen or reversed Hoogsteen hydrogen bonding, between complementary bases. Generally, nucleic acid consists of nitrogenous bases that are either pyrimidines (cytosine (C), uracil (U), and thymine (T)) or purines (adenine (A) and guanine (G)). These nitrogenous bases form hydrogen bonds between a pyrimidine and a purine, and the bonding of the pyrimidine to the purine is referred to as "base pairing." More specifically, A will hydrogen bond to T or U, and G will bond to C. "Complementary" refers to the base pairing that occurs between to distinct nucleic acid sequences or two distinct regions of the same nucleic acid sequence.

The terms specifically hybridizable and **specifically complementary** are terms that indicate a sufficient degree of complementarity such that stable and specific binding occurs between the oligonucleotide (or its analog) and the DNA or RNA target. The oligonucleotide or oligonucleotide analog need not be 100% complementary to its target sequence to be specifically hybridizable. An oligonucleotide or analog is specifically hybridizable when binding of the

oligonucleotide or analog to the target DNA or RNA molecule interferes with the normal function of the target DNA or RNA, and there is a sufficient degree of complementarity to avoid non-specific binding of the oligonucleotide or analog to non-target sequences under conditions where specific binding is desired, for example under physiological conditions in the case of *in vivo* assays or systems. Such binding is referred to as specific hybridization.

Hybridization conditions resulting in particular degrees of stringency will vary depending upon the nature of the hybridization method of choice and the composition and length of the hybridizing nucleic acid sequences. Generally, the temperature of hybridization and the ionic strength (especially the Na^+ concentration) of the hybridization buffer will determine the stringency of hybridization, though waste times also influence stringency. Calculations regarding hybridization conditions required for attaining particular degrees of stringency are discussed by Sambrook *et al.* (ed.), *Molecular Cloning: A Laboratory Manual*, 2nd ed., vol. 1-3, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989, chapters 9 and 11, herein incorporated by reference.

The following is an exemplary set of hybridization conditions and is not meant to be limiting.

Very High Stringency (detects sequences that share 90% sequence identity)

Hybridization:	5x SSC at 65°C for 16 hours
Wash twice:	2x SSC at room temperature (RT) for 15 minutes each
Wash twice:	0.5x SSC at 65°C for 20 minutes each

High Stringency (detects sequences that share 80% sequence identity or greater)

Hybridization:	5x-6x SSC at 65°C-70°C for 16-20 hours
Wash twice:	2x SSC at RT for 5-20 minutes each
Wash twice:	1x SSC at 55°C-70°C for 30 minutes each

Low Stringency (detects sequences that share greater than 50% sequence identity)

Hybridization:	6x SSC at RT to 55°C for 16-20 hours
Wash at least twice:	2x-3x SSC at RT to 55°C for 20-30 minutes each.

The hybridization conditions described above are adapted to hybridization, under high stringency conditions, of a molecule of nucleic acid of varying length from 20 nucleotides to several hundreds of nucleotides. The hybridization conditions described above may be adjusted as a function of the length of the nucleic acid whose hybridization is sought or of the type of labeling chosen, according to techniques known to one skilled in the art. Suitable hybridization conditions may, for example, be adjusted according to the teaching contained in the manual by Hames and Higgins, *Gene*

Transcription: A practical approach Transcription and Translation, IRL/Oxford University Press, 1993.

Hypercholesterolemia, or high cholesterol, is a high level of cholesterol in the blood, which can cause an arterial plaque to form and accumulate, leading to blockages in the arteries (atherosclerosis), increasing the risk for heart attack, stroke, circulation problems, and death. Hypercholesterolemia is measured relative to normal ranges or cholesterol obtained from population based studies.

***In vitro* amplification** includes techniques that increase the number of copies of a nucleic acid molecule in a sample or specimen. An example of amplification is the polymerase chain reaction, in which a biological sample collected from a subject is contacted with a pair of oligonucleotide primers, under conditions that allow for the hybridization of the primers to nucleic acid template in the sample. The primers are extended under suitable conditions, dissociated from the template, and then re-annealed, extended, and dissociated to amplify the number of copies of the nucleic acid. The product of *in vitro* amplification may be characterized by electrophoresis, restriction endonuclease cleavage patterns, oligonucleotide hybridization or ligation, and/or nucleic acid sequencing, using standard techniques. Other examples of *in vitro* amplification techniques include strand displacement amplification (see U.S. Patent No. 5,744,311); transcription-free isothermal amplification (see U.S. Patent No. 6,033,881); repair chain reaction amplification (see WO 90/01069); ligase chain reaction amplification (see EP-A-320 308); gap filling ligase chain reaction amplification (see U.S. Patent No. 5,427,930); coupled ligase detection and PCR (see U.S. Patent No. 6,027,889); and NASBA™ RNA transcription-free amplification (see U.S. Patent No. 6,025,134).

Isolated for the purposes of the present disclosure designates that a biological material (nucleic acid or protein) has been substantially separated from or purified away from other biological components in the cell or the organism in which the compound naturally occurs *i.e.*, other chromosomal and extra-chromosomal DNA and RNA, proteins and organelles. For example, a polynucleotide present in the natural state in a plant or an animal is not isolated. The same nucleotide separated from the cell and adjacent nucleic acids in which it is naturally inserted in the genome of the plant or animal is considered "isolated."

Labeled: A biomolecule attached covalently or noncovalently to a detectable label or reporter molecule. Typical labels include radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989 and Ausubel *et al.*, *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1998. For example, ATP can be labeled in any one of its three phosphate groups with radioisotopes such as ^{32}P or ^{33}P, or in its sugar moiety with radioisotopes such as ^{35}S.

Mammal: This term includes both human and non-human mammals. Similarly, the term **subject** includes both human and veterinary subjects.

Modulator: An agent that increases or decreases (modulates) the activity of a protein as measured by the change in an experimental parameter. A modulator can be essentially any compound, such as a chemotherapeutic agent, a polypeptide, a hormone, a nucleic acid, a sugar, a lipid and the like.

Mutation: Any change of the DNA sequence within a gene or chromosome. In some instances, a mutation will alter a characteristic or trait (phenotype), but this is not always the case. Types of mutations include base substitution point mutations (*e.g.*, transitions or transversions), deletions, and insertions. Missense mutations are those that introduce a different amino acid into the sequence of the encoded protein; nonsense mutations are those that introduce a new stop codon. In the case of insertions or deletions, mutations can be in-frame (not changing the frame of the overall sequence) or frame shift mutations, which may result in the misreading of a large number of codons (and often leads to abnormal termination of the encoded product due to the presence of a stop codon in the alternative frame).

This term specifically encompasses variations that arise through somatic mutation, for instance those that are found only in disease cells, but not constitutionally, in a given individual. Examples of such somatically-acquired variations include the point mutations that frequently result in altered function of various genes that are involved in development of cancers. This term also encompasses DNA alterations that are present constitutionally, that alter the function of the encoded protein in a readily demonstrable manner, and that can be inherited by the children of an affected individual. In this respect, the term overlaps with "polymorphism," as defined below, but generally refers to the subset of constitutional alterations that have arisen within the past few generations in a kindred and that are not widely disseminated in a population group. In particular embodiments, the term is directed to those constitutional alterations that have major impact on the health of affected individuals.

Nucleotide: This term includes, but is not limited to, a monomer that includes a base linked to a sugar, such as a pyrimidine, purine or synthetic analogs thereof, or a base linked to an amino acid, as in a peptide nucleic acid (PNA). A nucleotide is one monomer in a **polynucleotide.** The term **nucleotide** designates both the natural nucleotides (A, T, G, C) as well as the modified nucleotides that comprise at least one modification such as (1) an analog of a purine, (2) an analog of a pyrimidine, or (3) an analogous sugar, examples of such modified nucleotides being described, for example, in the PCT application No. WO 95/04/064. For the purposes of the present disclosure, a first polynucleotide is considered as being **complementary** to a second polynucleotide when each base of the first nucleotide is paired with the complementary base of the second polynucleotide whose orientation is reversed. The complementary bases are A and T (or A and U), or C and G.

A **nucleotide sequence** refers to the sequence of bases in a polynucleotide or nucleic acid. A **nucleic acid** is a polymeric compound comprised of covalently linked subunits called nucleotides. Nucleic acid includes polyribonucleic acid (RNA) and polydeoxyribonucleic acid (DNA), both of which may be single-stranded or double-stranded. DNA includes cDNA, genomic DNA, synthetic

DNA, and semi-synthetic DNA. The sequence of nucleotides that encodes a protein is called the sense sequence or coding sequence.

A **nucleic acid molecule** refers to the phosphate ester polymeric form of ribonucleosides (adenosine, guanosine, uridine or cytidine; "RNA molecules") or deoxyribonucleosides (deoxyadenosine, deoxyguanosine, deoxythymidine, or deoxycytidine; "DNA molecules"), or any phosphoester analogs thereof, such as phosphorothioates and thioesters, in either single stranded form, or a double-stranded helix. Double stranded DNA-DNA, DNA-RNA and RNA-RNA helices are possible. The term nucleic acid molecule, and in particular DNA or RNA molecule, refers only to the primary and secondary structure of the molecule, and does not limit it to any particular tertiary form. Thus, this term includes double-stranded DNA found, *inter alia*, in linear or circular DNA molecules (*e.g.*, restriction fragments), plasmids, and chromosomes. In discussing the structure of particular double-stranded DNA molecules, sequences may be described herein according to the normal convention of giving only the sequence in the 5' to 3' direction along the non-transcribed strand of DNA (*i.e.*, the strand having a sequence homologous to the mRNA). A **recombinant DNA molecule** is a DNA molecule that has undergone a molecular biological manipulation

Oligonucleotide: A plurality of joined nucleotides joined by native phosphodiester bonds, between about 6 and about 300 nucleotides in length. An oligonucleotide analog refers to moieties that function similarly to oligonucleotides but have non-naturally occurring portions. For example, oligonucleotide analogs can contain non-naturally occurring portions, such as altered sugar moieties or inter-sugar linkages, such as a phosphorothioate oligodeoxynucleotide. Functional analogs of naturally occurring polynucleotides can bind to RNA or DNA, and include peptide nucleic acid (PNA) molecules.

Oligonucleotides can be labeled, *e.g.*, with ^{32}P-nucleotides or nucleotides to which a label, such as biotin, has been covalently conjugated. In one embodiment, a labeled oligonucleotide can be used as a probe to detect the presence of a nucleic acid encoding an ABCA8 polypeptide of the disclosure. In another embodiment, oligonucleotides (one or both of which may be labeled) can be used as *in vitro* amplification primers, either for cloning full lengths or fragments of *ABCA8* nucleic acids, or to detect the presence of nucleic acids encoding *ABCA8*. In a further embodiment, an oligonucleotide of the disclosure can form a triple helix with *ABCA8* DNA molecules.

Generally, oligonucleotides are prepared synthetically, in some embodiments on a nucleic acid synthesizer. Accordingly, oligonucleotides can be prepared with non-naturally occurring phosphoester analog bonds, such as thioester bonds, etc.

Particular oligonucleotides and oligonucleotide analogs can include linear sequences up to about 200 nucleotides in length, for example a sequence (such as DNA or RNA) that is at least 6 bases, for example at least 8, 10, 15, 20, 25, 30, 35, 40, 45, 50, 100 or even 200 bases long, or from about 6 to about 50 bases, for example about 10-25 bases, such as 12, 15 or 20 bases.

Operably linked: A first nucleic acid sequence is operably linked with a second nucleic acid sequence when the first nucleic acid sequence is placed in a functional relationship with the second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the

promoter affects the transcription or expression of the coding sequence. Generally, operably linked DNA sequences are contiguous and, where necessary to join two protein-coding regions, in the same reading frame.

Open reading frame: A series of nucleotide triplets (codons) coding for amino acids without any internal termination codons. These sequences are usually translatable into a peptide.

Ortholog: Two nucleic acid or amino acid sequences are orthologs of each other if they share a common ancestral sequence and diverged when a species carrying that ancestral sequence split into two species. Orthologous sequences are also homologous sequences.

Pharmaceutically acceptable vehicle, carrier or excipient: A diluents or filler that is pharmaceutically acceptable for a chosen method of administration, is sterile, and may be aqueous or oleaginous suspensions formulated using suitable dispersing or wetting agents and suspending agents. The particular pharmaceutically acceptable carrier and the ratio of active compound to carrier are determined by the solubility and chemical properties of the composition, the particular mode of administration, and standard pharmaceutical practice. The pharmaceutically acceptable carriers useful with the compositions provided herein are conventional. Martin, *Remington's Pharmaceutical Sciences*, published by Mack Publishing Co., Easton, PA, 19th Edition, 1995, describes compositions and formulations suitable for pharmaceutical delivery of compounds and molecules herein disclosed.

Any nucleic acid, polypeptide, vector, or host cell of the disclosure will in some embodiments be introduced *in vivo* in a pharmaceutically acceptable vehicle or excipient. The phrase "pharmaceutically acceptable" refers to molecular entities and compositions that are physiologically tolerable and do not typically produce an allergic or similar untoward reaction, such as gastric upset, dizziness and so forth, when administered to a subject. In particular embodiments, the term "pharmaceutically acceptable" means approved by a regulatory agency of the Federal or a state government or listed in the U.S. Pharmacopeia or other generally recognized pharmacopeia for use in animals, and more particularly in humans. The terms "**excipient**" or "**carrier**" refer to a diluent, adjuvant, or vehicle with which the compound is administered. Such pharmaceutical excipients or carriers can be sterile liquids, such as water and oils, including those of petroleum, animal, vegetable or synthetic origin, such as peanut oil, soybean oil, mineral oil, sesame oil and so forth. Water or aqueous solution saline solutions and aqueous dextrose and glycerol solutions are in some embodiments employed as excipients or carriers, particularly for injectable solutions.

The disclosure contemplates delivery of a vector that will express a therapeutically effective amount of ABCA8 polypeptide for gene therapy applications. The phrase "**therapeutically effective amount**" is used herein to mean an amount sufficient to reduce by at least about 15 percent, in some embodiments by at least 50 percent, in other embodiments by at least 75, 80, or 90 percent, and in still other embodiments entirely prevent, a clinically significant deficit in the activity, function and response of the host. Alternatively, a therapeutically effective amount is sufficient to cause an improvement in a clinically significant condition in the host.

Polypeptide: A polymer in which the monomers are amino acid residues that are joined together through amide bonds. When the amino acids are alpha-amino acids, either the L-optical isomer or the D-optical isomer can be used, the L-isomers being preferred. The term **polypeptide** or **protein** as used herein encompasses any amino acid sequence and includes modified sequences such as glycoproteins. The term polypeptide is specifically intended to cover naturally occurring proteins, as well as those that are recombinantly or synthetically produced.

The term **polypeptide fragment** refers to a portion of a polypeptide that exhibits at least one useful epitope. The phrase **functional fragments of a polypeptide** refers to all fragments of a polypeptide that retain an activity, or a measurable portion of an activity, of the polypeptide from which the fragment is derived. Fragments, for example, can vary in size from a polypeptide fragment as small as an epitope capable of binding an antibody molecule to a large polypeptide capable of participating in the characteristic induction or programming of phenotypic changes within a cell. An epitope is a region of a polypeptide capable of binding an immunoglobulin generated in response to contact with an antigen. Thus, smaller peptides containing the biological activity of insulin, or conservative variants of the insulin, are thus included as being of use.

The term **soluble** refers to a form of a polypeptide that is not inserted into a cell membrane.

Conservative amino acid substitution tables providing functionally similar amino acids are well known to one of ordinary skill in the art. The following six groups are examples of amino acids that are considered to be conservative substitutions for one another:

1) Alanine (A), Serine (S), Threonine (T);

2) Aspartic acid (D), Glutamic acid (E);

3) Asparagine (N), Glutamine (Q);

4) Arginine (R), Lysine (K);

5) Isoleucine (I), Leucine (L), Methionine (M), Valine (V); and

6) Phenylalanine (F), Tyrosine (Y), Tryptophan (W).

An equivalent amino acid according to the present disclosure will be understood to mean for example replacement of a residue in the L form by a residue in the D form or the replacement of a glutamic acid (E) by a pyro-glutamic acid according to techniques well known to persons skilled in the art. According to another aspect, two amino acids belonging to the same class, that is to say two uncharged polar, nonpolar, basic or acidic amino acids, are also considered as equivalent amino acids.

Variations in the cDNA sequence that result in amino acid changes, whether conservative or not, are usually minimized in order to preserve the functional and immunologic identity of the encoded protein. The immunologic identity of the protein may be assessed by determining whether it is recognized by an antibody; a variant that is recognized by such an antibody is immunologically conserved. Any cDNA sequence variant will preferably introduce no more than twenty, and preferably fewer than ten amino acid substitutions into the encoded polypeptide. Variant amino acid sequences may, for example, be 80, 90 or even 95% or 98% identical to the native amino acid

sequence. Programs and algorithms for determining percentage identity can be found at the NCBI website.

A **variant polypeptide** includes any analogue, fragment, derivative, or mutant that is derived from a polypeptide or protein and that retains at least one biological property of the polypeptide or protein. Different variant polypeptides or variant proteins may exist in nature. These variants may be allelic variations characterized by differences in the nucleotide sequences of the structural gene coding for the protein, or may involve differential splicing or post-translational modification. Variant polypeptides also include a related (homologous) protein having substantially the same biological activity, but that has been obtained from a different species. In various embodiments, a **variant polypeptide** may contain up to 3, up to 5, up to 10, up to 15, or up to 25 or more different, deleted or added amino acid residues.

The skilled artisan can produce variant polypeptides having single or multiple amino acid substitutions, deletions, additions, or replacements. These variants may include, *inter alia*: (a) variants in which one or more amino acid residues are substituted with conservative or non-conservative amino acids, (b) variants in which one or more amino acids are added to the polypeptide or protein, (c) variants in which one or more of the amino acids includes a substituent group, and (d) variants in which the polypeptide or protein is fused with another polypeptide such as serum albumin, and a combination of these changes. The techniques for obtaining these variant polypeptides, including genetic (suppressions, deletions, mutations, etc.), chemical, and enzymatic techniques, are known to persons having ordinary skill in the art.

If such allelic variations, analogues, fragments, derivatives, mutants, and modifications, including alternative mRNA splicing forms and alternative post-translational modification forms, result in derivatives of the polypeptide that retain any of the biological properties of the polypeptide, such derivatives are included within the scope of this disclosure.

Polymorphism: Variant in a sequence of a gene, usually carried from one generation to another in a population. Polymorphisms can be those variations (nucleotide sequence differences) that, while having a different nucleotide sequence, produce functionally equivalent gene products, such as those variations generally found between individuals, different ethnic groups, geographic locations. The term polymorphism also encompasses variations that produce gene products with altered function, *i.e.*, variants in the gene sequence that lead to gene products that are not functionally equivalent. This term also encompasses variations that produce no gene product, an inactive gene product, or increased or increased activity gene product.

Polymorphisms can be referred to, for instance, by the nucleotide position at which the variation exists, by the change in amino acid sequence caused by the nucleotide variation, or by a change in some other characteristic of the nucleic acid molecule or protein that is linked to the variation (*e.g.*, an alteration of a secondary structure such as a stem-loop, or an alteration of the binding affinity of the nucleic acid for associated molecules, such as polymerases, RNases, and so forth).

Probes and primers: Nucleic acid probes and primers can be readily prepared based on the nucleic acid molecules provided in this disclosure. A probe comprises an isolated nucleic acid attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, enzyme substrates, co-factors, ligands, chemiluminescent or fluorescent agents, haptens, and enzymes. Methods for labeling and guidance in the choice of labels appropriate for various purposes are discussed, *e.g.*, in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989) and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992).

Primers are short nucleic acid molecules, preferably DNA oligonucleotides 10 nucleotides or more in length. More preferably, longer DNA oligonucleotides can be about 15, 17, 20, or 23 nucleotides or more in length. Primers can be annealed to a complementary target DNA strand by nucleic acid hybridization to form a hybrid between the primer and the target DNA strand, and then the primer extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification of a nucleic acid sequence, *e.g.*, by the polymerase chain reaction (PCR) or other nucleic-acid amplification methods known in the art.

Methods for preparing and using probes and primers are described, for example, in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989), Ausubel *et al.* (In *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1998), and Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, © 1991, Whitehead Institute for Biomedical Research, Cambridge, MA). One of ordinary skill in the art will appreciate that the specificity of a particular probe or primer increases with its length. Thus, for example, a primer comprising 30 consecutive nucleotides of ABCA8 encoding nucleotide will anneal to a target sequence, such as an ABCA8 gene homolog from the gene family contained within a human genomic DNA library, with a higher specificity than a corresponding primer of only 15 nucleotides. Thus, in order to obtain greater specificity, probes and primers can be selected that comprise at least 17, 20, 23, 25, 30, 35, 40, 45, 50 or more consecutive nucleotides of ABCA8 nucleotide sequences.

The disclosure thus includes isolated nucleic acid molecules that comprise specified lengths of the disclosed ABCA8 cDNA sequences. Such molecules can comprise at least 17, 20, 23, 25, 30, 35, 40, 45 or 50 or more consecutive nucleotides of these sequences, and can be obtained from any region of the disclosed sequences. By way of example, the ABCA8 cDNA sequences can be apportioned into halves, thirds or quarters based on sequence length, and the isolated nucleic acid molecules can be derived from the first or second halves of the molecules, from any of the three thirds or any of the four quarters. By way of example, the human ABCA8 cDNA, ORF, coding sequence and gene sequences can be apportioned into about halves, thirds or quarters based on sequence length, and the isolated nucleic acid molecules (*e.g.*, oligonucleotides) can be derived from the first or second halves of the molecules, from any of the three thirds, or any of the four quarters. The cDNA also could be divided into smaller regions, *e.g.* about eighths, sixteenths, twentieths, fiftieths and so forth,

with similar effect.

Another mode of division, provided by way of example, is to divide an ABCA8 encoding sequence based on the regions of the sequence that are relatively more or less homologous to other members of the ABC transporter family. Thus, nucleic acid molecules, for instance to be used as hybridization probe molecules, may be selected from the region encoding the N-terminal nucleotide binding domain (NBD1) region (*e.g.*, about residues 502-656, or a fragment thereof), or from the region encoding the C-terminal nucleotide binding domain (NBD2) (*e.g.*, about residues 1308-1621 or a fragment thereof), or the or the region encoding the N-terminal transmembrane domain (TM1) region (*e.g.*, about residues 1-501, or a fragment thereof) or the region encoding the C-terminal transmembrane domain (TM2) region (*e.g.*, about residues 656-1307) of the amino acid sequence encoding the human ABCA8 cDNA shown in SEQ ID NO: 1 (SEQ ID NO: 2).

Another mode of division is to select the 5' (upstream) and/or 3' (downstream) region associated with an ABCA8 gene, or to select an intron or portion thereof.

Promoter: A binding site in a DNA chain at which RNA polymerase binds to initiate transcription of messenger RNA by one or more nearby structural genes. A promoter contains the necessary elements for the transcription and translation of a protein-coding sequence. An expression vector containing the necessary elements for the transcription and translation of a protein-coding sequence may function as a promoter. The nucleotide sequence coding for the ABCA8 polypeptide or antigenic fragment, derivatives or analogs thereof, or functionally active derivatives, including chimeric proteins thereof, may be inserted into an expression vector containing the elements of a promoter, including a "promoter sequence," *i.e.*, a DNA regulatory region capable of binding RNA polymerase in a cell and initiating transcription of a downstream (3' direction) coding sequence. Thus, nucleic acids encoding the ABCA8 polypeptide of the disclosure can be operationally associated with a promoter in an expression vector of the disclosure. Both cDNA and genomic sequences can be cloned and expressed under control of such regulatory sequences. An expression vector also includes a replication origin.

Purified: In a more pure form than is found in nature. The term purified does not require absolute purity; rather, it is intended as a relative term. Thus, for example, a purified protein preparation is one in which the protein referred to is more pure than the protein in its natural environment within a cell.

The term **substantially purified** as used herein refers to a molecule (*e.g.*, a nucleic acid, polypeptide, oligonucleotide, etc.) that is substantially free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In one embodiment, the molecule is a polypeptide that is at least 50% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In another embodiment, the polypeptide is at least at least 80% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated. In yet other embodiments, the polypeptide is at least 90% or at least 95% free of other proteins, lipids, carbohydrates or other materials with which it is naturally associated.

Recombinant: A nucleic acid that has a sequence that is not naturally occurring or has a

sequence that is made by an artificial combination of two otherwise separated segments of sequence. This artificial combination can be accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, *e.g.*, by genetic engineering techniques.

Regulatory region means a nucleic acid sequence that regulates the expression of a nucleic acid. A regulatory region may include sequences that are naturally responsible for expressing a particular nucleic acid (a homologous region) or may include sequences of a different origin (a heterologous or recombinant region, responsible for expressing different proteins or even synthetic proteins). In particular, the regulatory sequences can be sequences of eukaryotic or viral genes or derived sequences that stimulate or repress transcription of a gene in a specific or non-specific manner, or in an inducible or non-inducible manner. Regulatory regions include origins of replication, RNA splice sites, enhancers, transcriptional termination sequences, signal sequences that direct the polypeptide into the secretory pathways of the target cell, and promoters.

A regulatory region from a **heterologous source** is a regulatory region that is not naturally associated with the expressed nucleic acid. Included among the heterologous regulatory regions are regulatory regions from a different species, regulatory regions from a different gene, hybrid regulatory sequences, and regulatory sequences that do not occur in nature, but that are designed for instance by an ordinarily skilled artisan.

Selection marker: A selection marker or **detectable marker** is a biochemical characteristic or function of a molecule that allows it to be identified based upon its performance, *e.g.*, β-galactosidase activity, thymidine kinase activity, resistance to antibiotics, transformation phenotype, occlusion body formation in baculovirus, etc.). A recombinant vector/host system can be identified and selected based upon the presence or absence of certain selection marker functions, caused by the insertion of foreign genes in the vector. For instance, if a nucleic acid encoding the ABCA8 polypeptide is inserted within the selection marker sequence of the vector, recombinants containing ABCA8 nucleic acid inserts can be identified by the absence of marker function (*e.g.*, loss of antibiotic resistance).

Sequence identity: The similarity between two nucleic acid sequences, or two amino acid sequences, is expressed in terms of the similarity between the sequences, otherwise referred to as sequence identity. Sequence identity is frequently measured in terms of percentage identity (or similarity or homology); the higher the percentage, the more similar the two sequences are. Homologs or orthologs of the ABCA8 protein, and the corresponding cDNA sequence, will possess a relatively high degree of sequence identity when aligned using standard methods. This homology will be more significant when the orthologous proteins or cDNAs are derived from species that are more closely related (*e.g.*, human and chimpanzee sequences), compared to species more distantly related (*e.g.*, human and *C. elegans* sequences).

Typically, ABCA8 orthologs are at least 90% identical at the nucleotide level and at least 98% identical at the amino acid level when comparing ABCA8 to an orthologous ABCA8.

Methods of alignment of sequences for comparison are well known in the art. Various

programs and alignment algorithms are described in: Smith and Waterman J. Mol. Biol. 147(1): 195-197, 1981; Needleman and Wunsch *J. Mol. Biol.* 48: 443-453, 1970; Pearson and Lipman *Proc. Natl. Acad. Sci. USA* 85: 2444-2448, 1988; Higgins and Sharp *Gene,* 73: 237-244, 1988; Higgins and Sharp *CABIOS* 5: 151-153, 1989; Corpet *et al. Nuc. Acids Res.* 16, 10881-10890, 1988; Huang *et al. Computer Appls. in the Biosciences* 8, 155-165, 1992; and Pearson *et al. Meth. Mol. Bio.* 24: 307-331, 1994. Furthermore, Altschul *et al.* (*J. Mol. Biol.* 215:403-410, 1990) present a detailed consideration of sequence alignment methods and homology calculations.

The NCBI Basic Local Alignment Search Tool (BLAST) (Altschul *et al. J. Mol. Biol.* 215: 403-410, 1990) is available from several sources, including the National Center for Biotechnology Information (NCBI, Bethesda, MD) and on the Internet, for use in connection with the sequence analysis programs blastp, blastn, blastx, tblastn and tblastx. The Search Tool can be accessed at the NCBI website, together with a description of how to determine sequence identity using this program.

An alternative indication that two nucleic acid molecules are closely related is that the two molecules hybridize to each other under stringent conditions. Hybridization conditions such as those described above are useful with the nucleic acid molecules of the disclosure.

Nucleic acid sequences that do not show a high degree of identity can nevertheless encode similar amino acid sequences, due to the degeneracy of the genetic code. It is understood that changes in nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid molecules that all encode substantially the same protein.

Specific binding agent: An agent that binds substantially only to a defined target. Thus an ABCA8 protein-specific binding agent binds substantially only an ABCA8 protein. As used herein, the phrase ABCA8 **protein-specific binding agent** includes anti-ABCA8 protein antibodies and other agents (such as soluble receptors) that bind substantially only to an ABCA8 protein, such as the ABCA8 proteins of the disclosure, or conservative variants thereof.

Anti-ABCA8 protein antibodies can be produced using standard procedures described in a number of texts, including Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988). The determination that a particular agent binds substantially only to the ABCA8 protein can readily be made by using or adapting routine procedures. One suitable *in vitro* assay makes use of the Western blotting procedure (described in many standard texts, including Harlow and Lane, *Antibodies, A Laboratory Manual*, CSHL, New York, 1988). Western blotting can be used to determine that a given ABCA8 protein binding agent, such as an anti-ABCA8 protein monoclonal antibody, binds substantially only to the ABCA8 protein.

A phosphospecific binding agent specifically binds to a peptide containing a phosphorylated residue. Shorter fragments of antibodies can also serve as specific binding agents. For instance, Fabs, Fvs, and single-chain Fvs (SCFvs) that bind to ABCA8 would be ABCA8-specific binding agents. These antibody fragments are defined as follows: (1) Fab, the fragment which contains a monovalent antigen-binding fragment of an antibody molecule produced by digestion of whole antibody with the enzyme papain to yield an intact light chain and a portion of one heavy chain; (2) Fab', the fragment of

an antibody molecule obtained by treating whole antibody with pepsin, followed by reduction, to yield an intact light chain and a portion of the heavy chain; two Fab' fragments are obtained per antibody molecule; (3) (Fab')2, the fragment of the antibody obtained by treating whole antibody with the enzyme pepsin without subsequent reduction; (4) F(ab')2, a dimer of two Fab' fragments held together by two disulfide bonds; (5) Fv, a genetically engineered fragment containing the variable region of the light chain and the variable region of the heavy chain expressed as two chains; and (6) single chain antibody (SCA), a genetically engineered molecule containing the variable region of the light chain, the variable region of the heavy chain, linked by a suitable polypeptide linker as a genetically fused single chain molecule. Methods of making these fragments are routine.

Subject: Living multi-cellular vertebrate organisms, a category that includes both human and non-human mammals.

Target sequence: "Target sequence" is a portion of ssDNA, dsDNA or RNA that, upon hybridization to a therapeutically effective oligonucleotide or oligonucleotide analog, results in the inhibition of expression. For example, hybridization of therapeutically effectively oligonucleotide to an ABCA8 target sequence results in inhibition of ABCA8 expression. Either an antisense or a sense molecule can be used to target a portion of dsDNA, as both will interfere with the expression of that portion of the dsDNA. The antisense molecule can bind to the plus strand, and the sense molecule can bind to the minus strand. Thus, target sequences can be ssDNA, dsDNA, and RNA.

Transfected: A process by which a nucleic acid molecule is introduced into a cell, for instance by molecular biology techniques, resulting in a **transfected cell.** As used herein, the term transfection encompasses all techniques by which a nucleic acid molecule might be introduced into such a cell, including transduction with viral vectors, transfection with plasmid vectors, and introduction of DNA by electroporation, lipofection, and particle gun acceleration. A cell has been **transformed** by exogenous or heterologous DNA when the transfected DNA effects a phenotypic change. The transforming DNA may be integrated (covalently linked) into chromosomal DNA making up the genome of the cell.

Treating a disease: Includes inhibiting or preventing the partial or full development or progression of a disease, for example in a person who is known to have a predisposition to a disease. An example of a person with a known predisposition is someone with a history of diabetes in his or her family, or who has been exposed to factors that predispose the subject to a condition, such as lupus or rheumatoid arthritis. Furthermore, treating a disease refers to a therapeutic intervention that ameliorates at least one sign or symptom of a disease or pathological condition, or interferes with a pathophysiological process, after the disease or pathological condition has begun to develop.

Variant nucleic acids according to the disclosure are understood to mean a nucleic acid that differs by one or more bases relative to the reference polynucleotide. A variant nucleic acid may be of natural origin, such as an allelic variant that exists naturally, or it may be a non-natural variant obtained, for example, by mutagenic techniques.

In general, the differences between the reference (for instance, wild-type) nucleic acid and the variant nucleic acid are small such that the nucleotide sequences of the reference nucleic acid and of the variant nucleic acid are very similar and, in many regions, identical. The nucleotide modifications present in a variant nucleic acid may be silent, which means that they do not alter the amino acid sequences encoded by the variant nucleic acid. In some embodiments, nucleic acid variants will contain up to 5 different residues compared to the reference sequence, in other embodiments up to 10, up to 25, up to 50 or more.

However, the changes in nucleotides in a variant nucleic acid may also result in substitutions, additions or deletions in the polypeptide encoded by the variant nucleic acid in relation to the polypeptides encoded by the reference nucleic acid. In addition, nucleotide modifications in the coding regions may produce conservative or non-conservative substitutions in the amino acid sequence of the polypeptide. Such variant polypeptides in some embodiments will contain no more than 3 amino acid changes, no more than 5, no more than 10, no more than 15, or in some instances 25 or more differences than a reference sequence.

In some embodiments, the variant nucleic acids according to the disclosure encode polypeptides that substantially conserve the same function or biological activity as the polypeptide of the reference nucleic acid or alternatively the capacity to be recognized by antibodies directed against the polypeptides encoded by the initial reference nucleic acid.

Some variant nucleic acids will thus encode mutated forms of the polypeptides whose systematic study will make it possible to deduce structure-activity relationships of the protein in question. Knowledge of these variants in relation to the disease studied is important as it helps to elucidate the molecular cause of the pathology.

Vector: A nucleic acid molecule as introduced into a host cell, thereby producing a transfected host cell. Recombinant DNA vectors are vectors having recombinant DNA. A vector can include nucleic acid sequences that permit it to replicate in a host cell, such as an origin of replication. A vector can also include one or more selectable marker genes and other genetic elements known in the art. Viral vectors are recombinant DNA vectors having at least some nucleic acid sequences derived from one or more viruses.

A polynucleotide may be included in a vector and/or such a polynucleotide may be included in a composition and remains nevertheless in the isolated state because of the fact that the vector or the composition does not constitute the polynucleotide's natural environment.

Unless otherwise explained, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. The singular terms "a," "an," and "the" include plural referents unless context clearly indicates otherwise. Similarly, the word "or" is intended to include "and" unless the context clearly indicates otherwise. "Comprises" means "includes." It is further to be understood that all base sizes or amino acid sizes, and all molecular weight or molecular mass values, given for nucleic acids or polypeptides are approximate, and are provided for description. Although methods and materials similar or

equivalent to those described herein can be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In case of conflict, the present specification, including explanations of terms, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

III. Description of Several Specific Embodiments

Provided herein is the amino acid sequence encoding ABCA8, as SEQ ID NO: 2. Specific embodiments include purified proteins having high identity (*e.g.*, 98%) to SEQ ID NO: 2. The disclosure includes oligonucleotides having high identity (*e.g.*, 80%, 90%) to residues 597-636 of SEQ ID NO: 2. The provided ABCA8 proteins have ABCA8 protein biological activity, for instance in that they can complement an *ABCA8* null phenotype. Specific embodiments include ABCA8 proteins having biological activity as extra- and/or intracellular membrane transport proteins that can complement *ABCA8* null phenotypes by ameliorating the respective transport deficiencies.

Also provided is the nucleic acid sequence encoding *ABCA8*, as SEQ ID NO: 1. Recombinant polynucleotides encoding this sequence are provided herein. Nucleic acid molecules that hybridize to the sequences disclosed herein are provided.

Specific binding agents that bind to epitopes of the ABCA8 protein and to specific regions within the protein (*e.g.*, residues 597-636 of SEQ ID NO: 2) are provided.

Methods to diagnose and detect defects or alterations in ABCA8 expression are provided, as are methods for screening for specific binding agents of ABCA8.

Transformed cells, recombinant vectors and transgenic animals containing the disclosed sequences are provided.

In specific embodiments, the methods are used to detect hypercholesterolemia, drug resistance, retinal degeneration, defective intra- or extracellular transport of ions, amino acids, peptides, sugars, vitamins or steroid hormones, or neurological disease. Other methods are used to detect chemotherapy resistant cells, or employ primers having sequences identical to at least 10 contiguous nucleotides of disclosed sequences, including SEQ ID NOS: 1, 81, 20, 33, 48, 61 or 77.

Antibodies specific to ABCA8, and the use of such antibodies (*e.g.*, in Western blot to ELISA assays) are disclosed.

Kits for using the ABCA8 protein are disclosed, including kits for screening defects in ABCA8 biological activity, and kits used to assay particular defects in individuals with defective ABCA8, including altered extra- or intracellular transport. Components useful in the provided kits are disclosed, including nucleotide probes, nucleic acid molecules, reagents of use, and instructions for using the kits.

Pharmaceutical compositions including polypeptides encoding the ABCA8 protein are provided.

IV. Human ABCA8 Protein and Nucleic Acid Sequences

Provided herein is a novel human ABCA-family gene, designated *ABCA8*. The new gene is a member of an ABCA gene cluster that is closely spaced and arranged head-to-tail at a single region of chromosome 17q24, all five members of this cluster encode putative full transporters (see Figure 1). The *ABCA8* gene encodes a protein proposed to be involved in the energy-dependent transport of one or a variety of substances, for example ions, metals, amino acids, lipids, peptides, sugars, vitamins and steroid hormones. The newly discovered gene also shows considerable conservation of the amino acid sequences compared to other known ABCA genes, particularly within the transmembrane region (TM) and the ATP-binding regions (NBF), and has a similar gene organization to *ABCA1*.

The full-length cDNA for human *ABCA8* (SEQ ID NO: 1) is 5797 base pairs long, which is somewhat longer than the *ABCA8* sequence previously released in the GenBank database (Accession # NM_007168, corresponding to putative protein with Accession # NP_09099.1). The ABCA8 ORF encodes a protein of 1621 amino acids, having a predicted molecular weight of approximately 184 kDa and a predicted pI of about 7.11.

The human *ABCA8* cDNA as identified herein includes a region of 120 residues, corresponding to nucleotides 1927-2046 of SEQ ID NO: 1 and encoding the amino acids at residues 597-636 of SEQ ID NO: 2, that were previously completely unknown. This newly identified region overlaps the first Nucleotide Binding Fold (NBF1) of ABCA8, and it is believed that the protein previously available in GenBank (Accession # NP_009099.1) is non-functional or dis-functional for ABC transporter activity due to the lack of the complete NBF1.

As provided herein, the human ABCA8 protein comprises the deduced protein sequence (SEQ ID NO: 2) corresponding to the herein described human *ABCA8* cDNA sequence (SEQ ID NO: 1). This protein sequence shows high sequence homology to other members of the ABCA family. Figure 2 shows an alignment of the provided ABCA8 sequence (referred to in the figure as *huest990006*) and other members of the ABCA gene family.

The ABCA8 protein contains the following domains: (1) an N-terminal transmembrane domain (TM1) that falls within residues 1-501; (2) a Nucleotide Binding Domain (NBF1) that falls within residues 502-656; (3) a C-terminal transmembrane domain (TM2) that falls within residues 656-1307; (2) a second Nucleotide Binding Domain (NBF2) that falls within residues 1308-1621.

ABCA8 is a single-copy gene that is highly expressed in ovarian tissue. The homology between the protein functional domains (e.g. transmembrane and nucleotide binding regions) of *ABCA8* to other ABCA family members indicate that the protein product of ABCA8 is likely to be important in extra- and intracellular transport, as has been shown with *ABCA1*.

In the present instance, ABCA8 maps to mouse chromosome 11, 64-67 cM from the centromere.

ABCA8 cDNA sequences were used to pinpoint the location of *ABCA8* in the human

genome. *ABCA8* was localized by amplification of monochrome hybrids and radiation hybrids using the Polymerase Chain Reaction (PCR) (see Morten *et al.*, *Hum. Genet.* 88(2): 200-203, 1991). Subsequently, the position of *ABCA8* on the draft human genome map was determined using chromosomal assignment and somatic cell hybrids (CASH) (see Ryu *et al.*, *Mol Cells* 10(5): 598-600, 2000).

RT-PCR analysis shows that ABCA8 is expressed in heart, ovary, testes and liver. However, in a recent study, Tsuruoka *et al.*, found that Northern Blot analysis revealed that expression is found in most organs, and in the highest levels in hearth, skeletal muscle, and liver (see Tsuruoka *et al.*, *Biochem. Biophys Res. Commun.* 298(1):41, 2002).

ABCA8 maps to mouse chromosome 11, 64-67 cM from the centromere. *ABCA8* is therefore a positional candidate for pathologies linked to this chromosome.

This disclosure provides ABCA8 transporter proteins and variants thereof, and nucleic acid molecules encoding these proteins, including cDNA sequences. In specific embodiments, these sequences are used for ameliorating, treating, detecting, prognosing, and diagnosing diseases and conditions believed to be associated with altered *ABCA8* expression (based on homology to other ABCA family members), such as hypercholesterolemia, drug resistance, retinal degeneration, resistance to chemotherapy, or neurological disease. The prototypical *ABCA8* sequences (*e.g.* SEQ ID NOS: 1 and 2) are human, and the use of these sequences to produce transgenic animals having increased or decreased levels of ABCA8 protein is provided, as are diagnostic methods to detect defects or alterations in *ABCA8* expression or ABCA8 protein production.

The human *ABCA8* cDNA as identified herein includes a region of 120 residues, corresponding to nucleotides 1927-2046 of SEQ ID NO: 1 and encoding the amino acids at residues 597-636 of SEQ ID NO: 2, that were previously completely unknown. This newly identified region overlaps the first Nucleotide Binding Fold (NBF1) of ABCA8, and it is believed that the protein previously available in GenBank (Accession # NP_009099.1) is non-functional or dis-functional for ABC transporter activity due to the lack of the complete NBF1.

In some embodiments, provided ABCA8 proteins have ABCA8 protein biological activity, for instance in that they can complement an ABCA8 null phenotype by ameliorating the associated transport deficiencies. One specific embodiment is thus an isolated human ABCA8 protein predicted to have an estimated molecular weight of about 184 kDa, wherein the human ABCA8 protein comprises the amino acid sequence as set forth in SEQ ID NO: 2 (or a conservative variant thereof), the protein is associated with extra- and/or intracellular membrane transport, and it can complement an *ABCA8* null phenotype by ameliorating the transport deficienc(ies).

In another embodiment, an *ABCA8* nucleic acid is provided that has altered expression (*e.g.*, increased or decreased expression, such as altered transcription of *ABCA8* mRNA, a mutated or deleted expression product, improper subcellular localization of a protein, etc.) as compared to a control nucleic acid (*e.g.*, a nucleic acid amplified, using positive control sequences, from a subject not suffering from the biological condition). In some embodiments, expression of an *ABCA8* nucleic

acid is more than 50%, more than 75%, more than 100%, more than 200%, or more than 300% elevated or decreased when compared to a suitable control. Suitable controls include the expression level of *ABCA8* in a subject having normal transporter activity in a known sample, or a standard value as assigned by one of ordinary skill in the art as a suitable standard value.

It will be apparent to one of ordinary skill in the art that the cDNAs disclosed herein and sequences derived from these may be utilized in many applications, including but not limited to, studies of the expression of the *ABCA8* gene and mutants thereof, studies of the function of the ABCA8 protein, the generation of antibodies to the ABCA8 protein, and diagnosis and therapy of ABCA8 deleted or mutated in subjects to prevent or treat the defects in cell and tissue development, such as transporter deficiency. Descriptions of applications using of ABCA8 cDNA, or fragments thereof, are therefore intended to comprehend the use of the *ABCA8* nucleic acids and variants thereof.

It has been determined that *ABCA8* mRNA has a unique expression pattern compared to other ABCs, suggesting that the corresponding protein may perform tissue-specialized functions. In effect, expression patterns showed that the ABCA8 transcript is almost ubiquitous, with the strongest expression found in ovarian tissues. The following table summarizes the relative expression patterns of ABCA8 mRNA:

Table 1: Expression patterns of ABCA8 mRNA.

TISSUE	ABCA8
BRAIN	+
HEART	++
KIDNEY	+
LIVER	++
LUNG	+
PANCREAS	-
PLACENTA	+
SKELETAL MUSCLE	-
COLON	+
OVARY	+++
PERIPHERAL BLOOD LYMPHOCYTES	-
PROSTATE	+
SMALL INTESTINE	+
SPLEEN	+
TESTIS	++
THYMUS	+

The cDNA sequence *ABCA8* (SEQ ID NO: 1) consists of 5797 nucleotides and contains a 4863 nucleotide coding sequence corresponding to a 1621 amino acid (aa) ABCA8 polypeptide (SEQ ID NO: 2). The cDNA molecule of human *ABCA8* has the nucleotide sequence as set forth in SEQ ID NO: 1 and includes an open reading frame beginning from the nucleotide at position 139 (base A

of the ATG codon for initiation of translation) to the nucleotide at position 5004 (base A of the TGA stop codon).

The *ABCA8* cDNA composing SEQ ID NO: 1 encodes a full length ABCA8 polypeptide of 1621 amino acids corresponding to the amino acid sequence of SEQ ID NO: 2.

The ABCA8 protein also demonstrates high conservation with other members of the ABCA family, as set forth in Table 2 (Figure 2), as well as in particular conserved regions within these proteins (Tables 3-6).

Table 2: Amino Acid Homology / Identity percentages across the full length of members of the ABCA family.

Total Sequence	ABCA5	ABCA6	ABCA8	ABCA9	ABCA10	ABCA1
ABCA5	100 / 100					
ABCA6	52.9 / 42.8	100 / 100				
ABCA8	52.4 / 42.4	67 / 59.7	100 / 100			
ABCA9	52.6 / 42.7	67.4 / 59.4	78.2 / 71.6	100 / 100		
ABCA10	53.2 / 43.4	69.5 / 62.3	68.1 / 61.1	70.3 / 62.1	100 / 100	
ABCA1	41.5 / 30.8	42.8 / 31	42.8/32	41.1 / 30.9	41.2 / 30.6	100 / 100

Table 3: Homology / Identity percentages in the N terminal transmembrane domain (residues 1 to 501 of ABCA8).

TMN domain	ABCA5	ABCA6	ABCA8	ABCA9	ABCA10	ABCA1
ABCA5	100 / 100					
ABCA6	47 / 34.2	100 / 100				
ABCA8	46.5 / 35	70.2 / 59.1	100 / 100			
ABCA9	46.3 / 37.8	64.2 / 55.7	76.9 / 68.5	100 / 100		
ABCA10	43.4 / 32.3	68.5 / 60.4	70.7 / 60.8	65.5 / 57.9	100 / 100	
ABCA1	36.5 / 23.1	34.2 / 20	39.8 / 27.6	40.6 / 27.9	35.4 / 24.4	100 / 100

Table 4: Homology / Identity percentages in the C terminal transmembrane domain (residues 656-1307 of ABCA8).

TMC domain	ABCA5	ABCA6	ABCA8	ABCA9	ABCA10	ABCA1
ABCA5	100 / 100					
ABCA6	43.7 / 33.7	100 / 100				
ABCA8	48.2 / 31.8	53.8 / 44.2	100 / 100			
ABCA9	47.3 / 33.7	57.2 / 48.2	64.1 / 52.9	100 / 100		
ABCA10	47 / 35.4	57 / 47	54.3 / 43	57.4 / 44.4	100 / 100	
ABCA1	33 / 21.6	32 / 21.4	39 / 24.8	35.3 / 26.8	34.7 / 22.4	100 / 100

Table 5: Homology / Identity percentages in the Nucleotide Binding Domain 1 (NBF1) (residues 502-656 of ABCA8).

NBF1 domain	ABCA5	ABCA6	ABCA8	ABCA9	ABCA10	ABCA1
ABCA5	100 / 100					
ABCA6	70.2 / 60.3	100 / 100				
ABCA8	71.8 / 62.5	85.4 / 78.6	100 / 100			
ABCA9	65.5 / 58,2	80.2 / 72.8	88.5 / 81.8	100 / 100		
ABCA10	69.8 / 62.1	83.2 / 77,2	82.8 / 79.2	81.9 / 75.8	100 / 100	
ABCA1	56.8 / 48.5	53.5 / 43.5	61.2 / 50.5	51.7 / 43.8	56.5 / 45.6	100 / 100

Table 6: Homology / Identity percentages in the Nucleotide Binding Domain 2 (NBF2) (residues 1308-1621 of ABCA8).

NBF2 domain	ABCA5	ABCA6	ABCA8	ABCA9	ABCA10	ABCA1
ABCA5	100 / 100					
ABCA6	63 / 56.1	100 / 100				
ABCA8	66.9 / 58.4	78 / 73.5	100 / 100			
ABCA9	65.2 / 57.0	77.6 / 72.6	94.5 / 91.8	100 / 100		
ABCA10	63.7 / 56.2	74.9 / 71.2	81 / 77.4	82.3 / 77.8	100 / 100	
ABCA1	46.4 / 37.8	46.3 / 37.9	46.9 / 38	47.3 / 39	46.4 / 37.7	100 / 100

Phylogenetic analysis of the ATP-binding domains of these ABCA proteins demonstrate that the N- and C-terminal domains form separate branches (Figure 3). The C-terminal ATP-binding domains of the 17q24 genes are more closely related to the C-terminal domains of the other ABC1-like genes than to the N-terminal domains of the same proteins. Thus, the entire ABC1 subfamily appears to have arisen from a single ancestral full transporter gene. However, the genes in the 17q24 cluster form a distinct group within the ABC1 subfamily.

V. Production of ABCA8 Nucleic Acids

With the provision herein of the sequence of the full length ABCA8 protein (SEQ ID NO: 2) and cDNA (SEQ ID NO: 1), *in vitro* nucleic acid amplification [such as polymerase chain reaction (PCR)] may be utilized as a simple method for producing ABCA8 encoding sequences. The following provides representative techniques for preparing cDNA in this manner.

Total RNA is extracted from human cells by any one of a variety of methods well known to those of ordinary skill in the art (Sambrook *et al.*, in *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989; and Ausubel *et al.*, in *Current Protocols in Molecular Biology*, Greene Publ. Assoc. and Wiley-Intersciences, 1992 provide descriptions of methods for RNA isolation).

In one embodiment, primary cells are obtained from normal tissues. In another embodiment cells are obtained from tissues from subjects exhibiting the effects of ABCA8 transporter deficiency. In yet another embodiment, cell lines derived from normal or transporter-deficient tissues are used as a source of such RNA. The extracted RNA is then used, for example, as a template for performing reverse transcription (RT)-PCR amplification of cDNA. Methods and conditions for RT-PCR are described in Kawasaki *et al.*, (In *PCR Protocols, A Guide to Methods and Applications*, Innis *et al.*

(eds.), 21-27, Academic Press, Inc., San Diego, California, 1990).

The selection of amplification primers will be made according to the portion(s) of the cDNA that is to be amplified. In one embodiment, primers may be chosen to amplify a segment of a cDNA or, in another embodiment, the entire cDNA molecule. Variations in amplification conditions may be required to accommodate primers and amplicons of differing lengths and composition; such considerations are well known in the art and are discussed for instance in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). By way of example, the ABCA8 protein coding portion may be amplified using primers, such as SEQ ID NO: 5 and 6.

Similarly, the primers set forth as SEQ ID NO: 3 and 4 can be used to amplify SEQ ID NO: 1; the primers set forth as SEQ ID NO: 7 and 8 can be used to amplify a particular region (see Example 3) within ABCA8; and the primers set forth as SEQ ID NO: 9 and 10 can be used to amplify an ABCA8 encoding sequence. These primers are illustrative only; one skilled in the art will appreciate that many different primers may be derived from the provided cDNA or gene sequence in order to amplify particular regions of the *ABCA8* cDNA, as well as the complete sequence of human *ABCA8* cDNA (SEQ ID NO: 1).

Re-sequencing of PCR products obtained by these amplification procedures is recommended; this will facilitate confirmation of the amplified sequence and will also provide information on natural variation on this sequence in different populations or species. Oligonucleotides derived from the provided *ABCA8* sequences provided may be used in such sequencing methods.

Orthologs of human *ABCA8* can be cloned in a similar manner, where the starting material consists of cells taken from a non-human species. Orthologs will generally share at least 50% sequence homology with the disclosed human *ABCA8* cDNA. Where the non-human species is more closely related to humans, the sequence homology will in general be greater. Closely related orthologous *ABCA8* molecules may share at least 75%, at least 80%, at least 90%, at least 95%, or at least 98% sequence homology with the disclosed human sequences.

Oligonucleotides derived from the human *ABCA8* cDNA sequence (SEQ ID NO: 1) are encompassed within the scope of the present invention. Such oligonucleotide primers may comprise a sequence of at least 10 consecutive nucleotides of the *ABCA8* nucleic acid sequence. To enhance amplification specificity, oligonucleotide primers comprising at least 15, 25, 30, 35, 40, 45 or 50 or more consecutive nucleotides of these sequences may also be used. These primers for instance may be obtained from any region of the disclosed sequences. By way of example, the human *ABCA8* cDNA, ORF and gene sequences may be apportioned into about halves or quarters based on sequence length, and the isolated nucleic acid molecules (*e.g.*, oligonucleotides) may be derived from the first or second halves of the molecules, or any of the four quarters. The human *ABCA8* cDNA, shown in SEQ ID NO: 1, can be used to illustrate this example. The human *ABCA8* cDNA shown in SEQ ID NO: 1 is 5797 nucleotides in length and may be hypothetically divided into about halves (nucleotides

1-2899 and 2900-5797) or about quarters (nucleotides 1-1449, 1450-2899, 2900-4348 and 4349-5797). Another advantageous portion of the reported *ABCA8* locus is about residue 2026 to about residue 3113 of SEQ ID NO: 1, or the reverse complement of this sequence.

Nucleic acid molecules may be selected that comprise at least 10, 15, 20, 25, 30, 35, 40, 50 or 100 or more consecutive nucleotides of any of these or other portions of the human *ABCA8* cDNA, or of the 5' or 3' flanking regions.

VI. ABCA8 Sequence Variants

With the provision of human ABCA8 protein (SEQ ID NO: 2) and corresponding nucleic acid sequences (SEQ ID NO: 1) herein, the creation of variants of these sequences is now enabled.

In one embodiment, variant ABCA8 proteins include proteins that differ in amino acid sequence from the human ABCA8 sequences disclosed but that share at least 65% amino acid sequence identity with the provided human ABCA8 protein. In other embodiments, other variants will share at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% amino acid sequence identity. The *ABCA8* gene may be isolated by routine procedures, such as those provided in Example 1. For instance, the *ABCA8* gene may be isolated by homology screening using the cDNA sequence and the BLAST program. Direct sequencing, using the "long-distance sequence method," of one or more BAC or PAC clones that contain the ABCA8 sequence can be employed. Manipulation of the nucleotide sequence of ABCA8 using standard procedures, including in one specific, non-limiting, embodiment, site-directed mutagenesis or in another specific, non-limiting, embodiment, PCR, can be used to produce such variants. The simplest modifications involve the substitution of one or more amino acids for amino acids having similar biochemical properties. These so-called conservative substitutions are likely to have minimal impact on the activity of the resultant protein.

In another embodiment, more substantial changes in transporter function or other protein features may be obtained by selecting amino acid substitutions that are less conservative than conservative substitutions. In one specific, non-limiting, embodiment, such changes include changing residues that differ more significantly in their effect on maintaining polypeptide backbone structure (*e.g.*, sheet or helical conformation) near the substitution, charge or hydrophobicity of the molecule at the target site, or bulk of a specific side chain. The following specific, non-limiting, examples are generally expected to produce the greatest changes in protein properties: (a) a hydrophilic residue (*e.g.*, seryl or threonyl) is substituted for (or by) a hydrophobic residue (*e.g.*, leucyl, isoleucyl, phenylalanyl, or valyl); (b) a cysteine or proline is substituted for (or by) any other residue; (c) a residue having an electropositive side chain (*e.g.*, lysyl, arginyl, or histadyl) is substituted for (or by) an electronegative residue (*e.g.*, glutamyl or aspartyl); or (d) a residue having a bulky side chain (*e.g.*, phenylalanine) is substituted for (or by) one lacking a side chain (*e.g.*, glycine).

In other embodiments, changes in transporter activity or other protein features may be obtained by mutating, substituting or deleting regions of ABCA8 that have a known function, or

regions where the function is yet to be determined. For instance, a NBD (nucleotide binding domain) motif of ABCA8 (corresponding to about residues 502-656 or 1308-1621 of SEQ ID NO: 2) can be deleted, substituted with the NBD of another protein or a synthetic NBD, or residues within one or both of the NBD motif can be mutated. In other embodiments, residues within the two ABCA8 transmembrane domains (TM) (*e.g.*, residues 1-501 or 656-1307) are mutated or deleted, or the TM is substituted with a transmembrane domain of another protein or a synthetic TM.

In one specific, non-limiting, embodiment, a mutation is made within residues 597-636 of SEQ ID NO: 2.

Variant ABCA8 encoding sequences may be produced by standard DNA mutagenesis techniques. In one specific, non-limiting, embodiment, M13 primer mutagenesis is performed. Details of these techniques are provided in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Ch. 15., CSHL, New York, 1989). By the use of such techniques, variants may be created that differ in minor ways from the human ABCA8 sequences disclosed. In one embodiment, DNA molecules and nucleotide sequences that are derivatives of those specifically disclosed herein, and which differ from those disclosed by the deletion, addition, or substitution of nucleotides while still encoding a protein that has at least 90% sequence identity with the human ABCA8 encoding sequence disclosed (SEQ ID NO: 1), are comprehended by this disclosure. In some embodiments, at least one or more, at least 5 or more, at least 10 or more, at least 15 or more, at least 20 or more, or at least 25 or more nucleotides are deleted, added, or substituted while still encoding a protein that has at least 90% sequence identity with the human ABCA8 encoding sequence disclosed (SEQ ID NO: 1). In other embodiments, more closely related nucleic acid molecules that share at least 95%, or at least 98% nucleotide sequence identity with the disclosed ABCA8 sequences (*e.g.*, SEQ ID NO: 1 and 2) and nucleic acid molecule sharing at least 80% or more sequence identity with residues 597-636 of SEQ ID NO: 2 are comprehended by this disclosure. In one embodiment, such variants may differ from the disclosed sequences by alteration of the coding region to fit the codon usage bias of the particular organism into which the molecule is to be introduced.

In other embodiments, the coding region may be altered by taking advantage of the degeneracy of the genetic code to alter the coding sequence such that, while the nucleotide sequence is substantially altered, it nevertheless encodes a protein having an amino acid sequence substantially similar to the disclosed human ABCA8 protein sequences. For example, because of the degeneracy of the genetic code, four nucleotide codon triplets – (GCT, GCG, GCC and GCA) - code for alanine. The coding sequence of any specific alanine residue within the human ABCA8 protein, therefore, could be changed to any of these alternative codons without affecting the amino acid composition or characteristics of the encoded protein. Based upon the degeneracy of the genetic code, variant DNA molecules may be derived from the cDNA and gene sequences disclosed herein using standard DNA mutagenesis techniques as described above, or by synthesis of DNA sequences. Thus, this disclosure also encompasses nucleic acid sequences that encode an ABCA8 protein, but which vary from the disclosed nucleic acid sequences by virtue of the degeneracy of the genetic code.

In one embodiment, variants of ABCA8 may also be defined in terms of their sequence identity with the prototype human ABCA8 protein. As described above, human ABCA8 proteins share at least 65%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, or at least 98% amino acid sequence identity with the human ABCA8 protein (SEQ ID NO: 2). In some embodiments, at least one or more, at least 5 or more, at least 10 or more, at least 15 or more, at least 20 or more, or at least 25 or more amino acids are deleted, added, or substituted while still encoding a protein that has at least 90% sequence identity with the ABCA8 encoding sequence (SEQ ID NO: 2). Nucleic acid sequences that encode such proteins/fragments readily may be determined simply by applying the genetic code to the amino acid sequence of an ABCA8 protein or fragment, and such nucleic acid molecules may readily be produced by assembling oligonucleotides corresponding to portions of the sequence.

Nucleic acid molecules that are derived from the human ABCA8 cDNA nucleic acid sequences include molecules that hybridize under low stringency, high stringency, or very high stringency conditions to the disclosed prototypical ABCA8 nucleic acid molecules, and fragments thereof.

Human ABCA8 nucleic acid encoding molecules (*e.g.*, the cDNA shown in SEQ ID NO: 1, and nucleic acids comprising this sequence), and orthologs and homologs of this sequence, may be incorporated into transformation or expression vectors.

It will also be apparent to one skilled in the art that homologs of the *ABCA8* gene may now be cloned from other species, such as the rat or a monkey, by standard cloning methods. Such homologs will be useful in the production of animal models demonstrating the formation and progression of a variety of tumors. In one embodiment, such orthologous ABCA8 molecules will share at least 65% sequence identity with the human ABCA8 nucleic acid disclosed herein; and in other embodiments, more closely related orthologous sequences will share at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 98% sequence identity with this sequence. In another embodiment, orthologous ABCA8 molecules will share at least 65% sequence, at least 70%, at least 75%, at least 80%, at least 90%, at least 95%, or at least 98% sequence identity with residues 597-636 of SEQ ID NO: 2. In yet other embodiments, orthologous ABCA8 molecules share high (*e.g.*, at least 80% or more) homology with residues 597-636 of SEQ ID NO: 2 and at least 90% homology with the remainder of SEQ ID NO: 2.

VII. Expression of ABCA8 Protein

With the provision of the full-length human ABCA8 encoding sequence (SEQ ID NO: 1), the expression and purification of ABCA8 protein by standard laboratory techniques is now enabled. Purified human ABCA8 protein may be used for functional analyses, antibody production, diagnostics, and patient therapy.

For instance, the DNA sequence of the full-length *ABCA8* cDNA (SEQ ID NO: 1) can be manipulated in studies to understand the expression of the gene and the function of its product. In

other embodiments, partial or full-length cDNA sequences, which encode ABCA8 protein, may be ligated into bacterial expression vectors.

Methods for expressing large amounts of protein from a cloned gene introduced into *Escherichia coli* (*E. coli*) may be utilized for the purification, localization and functional analysis of proteins. By way of example, fusion proteins consisting of amino terminal peptides encoded by a portion of the *E. coli lacZ* or *trpE* gene linked to ABCA8 proteins may be used to prepare polyclonal and monoclonal antibodies against these proteins. Thereafter, these antibodies may be used in other embodiments to purify proteins by immunoaffinity chromatography, in diagnostic assays to quantitate the levels of protein and to localize proteins in tissues and individual cells by immunofluorescence. Such antibodies may be specific for epitope tags, which can be added to the expression construct for identification and/or purification purposes.

Intact native protein also may be produced in *E. coli* or other cell culture systems in large amounts for functional studies. Methods and plasmid vectors for producing fusion proteins and intact native proteins in bacteria are described in Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Ch. 17, CSHL, New York, 1989). Such fusion proteins may be made in large amounts, are easy to purify, and can be used to elicit antibody response. In one embodiment, native proteins can be produced in bacteria by placing a strong, regulated promoter and an efficient ribosome binding site upstream of the cloned gene. If low levels of protein are produced, additional steps may be taken to increase protein production; if high levels of protein are produced, purification is relatively easy. Suitable methods are presented in Sambrook *et al.* and are well known in the art. In one embodiment, proteins expressed at high levels are found in insoluble inclusion bodies. Methods for extracting proteins from these aggregates are described by Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual*, Ch. 17, CSHL, New York, 1989).

Vector systems suitable for the expression of *lacZ* fusion genes include the pUR series of vectors (Ruther and Muller-Hill, *EMBO J.* 2:1791-1794, 1983), pEX1-3 (Stanley and Luzio, *EMBO J.* 3(6):1429-1434, 1984) and pMR100 (Gray *et al., Proc. Natl. Acad. Sci. USA* 79(21):6598-6602, 1982). Vectors suitable for the production of intact native proteins include pKC30 (Belfort *et al., J. Biol. Chem.* 258(3):2045-2051, 1983), pKK177-3 (Amann and Brosius, *Gene* 40(2-3):183-190, 1985) and pET-3 (Studier and Moffatt, *J. Mol. Biol.* 189(1):113-130, 1986). In one embodiment, ABCA8 fusion proteins may be isolated from protein gels, lyophilized, ground into a powder and used as an antigen. In other embodiments, the DNA sequence can also be transferred from its existing context to other cloning vehicles, such as other plasmids, bacteriophages, cosmids, animal viruses and yeast artificial chromosomes (YACs) (Burke *et al., Science* 236:806-812, 1987). These vectors may then be introduced into a variety of hosts including, but not limited to, somatic cells, and simple or complex organisms, such as, but not limited to, bacteria, fungi (Timberlake and Marshall, *Science* 244:1313-1317, 1989), invertebrates, plants, and animals (Pursel *et al., Science* 244:1281-1288, 1989), which cells or organisms are rendered transgenic by the introduction of the heterologous ABCA8 cDNA.

For expression in mammalian cells, the cDNA sequence is ligated to heterologous promoters. In one specific, non-limiting embodiment it may be ligated to the simian virus (SV) 40 promoter in the pSV2 vector (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072-2076, 1981), and introduced into cells, such as monkey COS-1 cells (Gluzman, *Cell* 23:175-182, 1981), to achieve transient or long-term expression. In one embodiment, the stable integration of the chimeric gene construct may be maintained in mammalian cells by biochemical selection, such as neomycin (Southern and Berg, *J. Mol. Appl. Genet.* 1:327-341, 1982) and mycophenolic acid (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072-2076, 1981). In another embodiment, cell lines expressing native ABCA8 are created. In yet another embodiment, cell lines expressing a mutant ABCA8 are created.

DNA sequences can be manipulated with standard procedures such as restriction enzyme digestion, fill-in with DNA polymerase, deletion by exonuclease, extension by terminal deoxynucleotide transferase, ligation of synthetic or cloned DNA sequences, site-directed sequence-alteration via single-stranded bacteriophage intermediate or with the use of specific oligonucleotides in combination with nucleic acid amplification. These techniques are known to those of ordinary skill in the art.

The *ABCA8* cDNA sequence (or portions derived from it, such as a nucleic acid encoding a polypeptide including an amino acid sequence selected from SEQ ID NO: 2 or residues 597-636 of SEQ ID NO: 2) or a mini gene (a cDNA with an intron and its own promoter) may be introduced into eukaryotic expression vectors by conventional techniques. These vectors are designed to permit the transcription of the cDNA in eukaryotic cells by providing regulatory sequences that initiate and enhance the transcription of the cDNA and ensure its proper splicing and polyadenylation. Vectors containing the promoter and enhancer regions of the SV40 or long terminal repeat (LTR) of the Rous Sarcoma virus and polyadenylation and splicing signal from SV40 are readily available (Mulligan *et al., Proc. Natl. Acad. Sci. USA* 78:1078-2076, 1981; Gorman *et al., Proc. Natl. Acad. Sci USA* 78:6777-6781, 1982). The level of expression of the cDNA can be manipulated with this type of vector by using promoters that have different activities (for example, the baculovirus pAC373 can express cDNAs at high levels in *S. frugiperda* cells (Summers and Smith, in *Genetically Altered Viruses and the Environment*, Fields *et al.* (Eds.) 22:319-328, CSHL Press, Cold Spring Harbor, New York, 1985) or by using vectors that contain promoters amenable to modulation, for example, a recombinant adenoviral vector containing a nuclear lacZ gene driven by a human ventricular/slow muscle myosin light chain 1 promoter (Shi *et al., Hum Gene Ther.* 8(4):403-410, 1997). The expression of the cDNA can be monitored in the recipient cells 24 to 72 hours after induction (transient expression).

Some vectors contain selectable markers such as the *gpt* (Mulligan and Berg, *Proc. Natl. Acad. Sci. USA* 78:2072-2076, 1981) or *neomycin* (Southern and Berg, *J. Mol. Appl. Genet.* 1:327-341, 1982) bacterial genes. These selectable markers permit selection of transfected cells that exhibit stable, long-term expression of the vectors (and therefore the cDNA). By way of example, the vectors

can be maintained in the cells as episomal, freely replicating entities by using regulatory elements of viruses, such as papilloma (Sarver *et al., Mol. Cell Biol.* 1:486-496, 1981) or Epstein-Barr (Sugden *et al., Mol. Cell Biol.* 5:410-413, 1985). In another embodiment, one can produce cell lines that have integrated the vector into genomic DNA. Both of these types of cell lines produce the gene product on a continuous basis. Alternatively, one can produce cell lines that have amplified the number of copies of the vector (and therefore of the cDNA as well) to create cell lines that can produce high levels of the gene product (Alt *et al., J. Biol. Chem.* 253: 1357-1370, 1978).

The vectors may contain an internal ribosomal entry site (IRES) between the cDNA and a marker gene, such as neomycin or enhanced green fluorescent protein (EGFP). The IRES allows for the simultaneous expression of the two elements from a single isoform. Ribosomes bind the isoform at both the 5' end to translate the cDNA and at the IRES to translate, in one specific, non-limiting embodiment, the antibiotic resistance marker, or in another specific, non-limiting embodiment, the fluorescent marker. The bicistronic expression via the IRES sequence provides a high degree of correlation between the antibiotic resistance and stable expression of the cDNA. Alternatively, only cells expressing the cDNA will show green fluorescence. Thus, the use of expression vectors containing an IRES is an efficient way to select for cells expressing the cDNA of interest.

The transfer of DNA into eukaryotic, in particular human or other mammalian cells, is now a conventional technique. Recombinant expression vectors can be introduced into the recipient cells as pure DNA (transfection) by, for example, precipitation with calcium phosphate (Graham and vander Eb, *Virology* 52(2):456-467, 1973) or strontium phosphate (Brash *et al., Mol. Cell Biol.* 7(5):2031-2035, 1987), electroporation (Neumann *et al., EMBO J* 1(7):841-845, 1982), lipofection (Felgner *et al., Proc. Natl. Acad. Sci USA* 84(21):7413-7417, 1987), DEAE dextran (Schenborn *et al., Methods Mol. Biol.* 130:147-153, 2000), microinjection (Mueller *et al., Cell* 15(2):579-585, 1978), protoplast fusion (Schafner, *Proc. Natl. Acad. Sci. USA* 77:2163-2167, 1980), or gene guns (Nishitani *et al., Cancer Gene Ther.* 9(2):156-163, 2002). In another embodiment, the ABCA8 cDNA sequence, or fragments thereof, can be introduced by infection with virus vectors. Systems are developed that use, for example, retroviruses (Huszar, *et al.*, Proc. Natl. Acad. Sci. U. S. A. 82(24):8587-8591, 1985), adenoviruses (Schaack *et al., Virology.* 291(1):101-109, 2001), or Herpes virus (Spaete *et al., Cell* 30(1):295-304, 1982). Techniques of use in packaging long transcripts can be found in Kochanek *et al.* (*Proc. Natl. Acad. Sci. USA* 93:5731-5739, 1996), Parks *et al.* (*Proc. Natl. Acad. Sci. USA* 93:13565-13570, 1996) and Parks and Graham (*J. Virol.* 71:3293-3298, 1997). In yet another embodiment, ABCA8 encoding sequences can be delivered to target cells *in vitro* via non-infectious systems, for instance liposomes.

These eukaryotic expression systems can be used for studies of *ABCA8* encoding nucleic acids and mutant forms of these molecules. Regulatory elements located in the 5' region of the *ABCA8* gene on genomic clones can be isolated from human genomic DNA libraries using the information contained herein. In other embodiments, the eukaryotic expression systems also may be used to study the function of the normal complete protein, specific portions of the protein, or of

naturally occurring or artificially produced mutant proteins.

In several embodiments, using the above techniques, expression vectors containing the *ABCA8* gene sequence or cDNA (*e.g.*, SEQ ID NO: 1), or fragments or variants or mutants thereof (*e.g.*, nucleotides 1927-2046 of SEQ ID NO: 1), can be introduced into human cells, mammalian cells from other species or non-mammalian cells, as desired. The choice of cell is determined by the purpose of the treatment. For example, in one specific, non-limiting embodiment monkey COS cells (Gluzman, *Cell* 23:175-182, 1981) that produce high levels of the SV40 T antigen and permit the replication of vectors containing the SV40 origin of replication are used. In other embodiments, Chinese hamster ovary (CHO), mouse NIH 3T3 fibroblasts, or human fibroblasts or lymphoblasts are used.

According to one embodiment of the disclosure, a "naked" ABCA8 polynucleotide is placed under the control of appropriate regulatory sequences, and is introduced by local injection at the level of the chosen tissue, for example myocardial tissue, the "naked" polynucleotide being absorbed by the myocytes of this tissue.

Compositions for use *in vitro* and *in vivo* including "naked" polynucleotides are for example described in published international Application WO 95/11307 as well as in Tacson *et al.*, *Nature Medicine*, 2(8):888-892 (1996) and Huygen *et al.*, *Nature Medicine*, 2(8):893-898 (1996).

According to a specific embodiment, a composition is provided for the *in vivo* production of ABCA8 protein. This composition includes a polynucleotide encoding the ABCA8 polypeptide placed under the control of appropriate regulatory sequences, in solution in a physiologically acceptable vector.

The quantity of vector that is injected into a host organism varies according to the site of the injection. As a guide, there may be injected between about 0.1 and about 100 μg of polynucleotide encoding an ABCA8 protein or variant or gragment thereof into the body of an animal, for instance into a patient likely to develop a disease-linked ABCA8 deficiency.

Embodiments described herein thus encompass recombinant vectors that comprise all or part of an ABCA8 encoding sequence, such as the *ABCA8* gene or cDNAs or variants thereof, for expression in a suitable host. In one embodiment, an *ABCA8* DNA is operatively linked in a vector to an expression control sequence in the recombinant DNA molecule so that the ABCA8 polypeptide can be expressed. In another embodiment, the expression control sequence may be selected from the group consisting of sequences that control the expression of genes of prokaryotic or eukaryotic cells and their viruses and combinations thereof. The expression control sequence may be specifically selected from the group consisting of the *lac* system, the *trp* system, the *tac* system, the *trc* system, major operator and promoter regions of phage lambda, the control region of coat protein, the early and late promoters of SV40, promoters derived from polyoma, adenovirus, retrovirus, baculovirus and simian virus, the promoter for 3-phosphoglycerate kinase, the promoters of yeast acid phosphatase, the promoter of the yeast alpha-mating factors and combinations thereof.

In another embodiment, the host cell, which may be transfected with a vector, may be

selected from the group consisting of *E. coli*, *Pseudomonas*, *Bacillus subtilis*, *Bacillus stearothermophilus* or other bacilli; other bacteria; yeast; fungi; insect; mouse or other animal; or plant hosts; or human tissue cells.

VIII. Suppression of ABCA8 Protein Expression

A diseased condition is in some instances attributed to overexpression of a protein (*e.g.*, HER-2/neu is overexpressed in breast cancer, *see* Kaya *et al.*, *Pathol. Oncol. Res.* 7(4): 279-283, 2001). In such instances, reversal of the over-expression (*e.g.*, suppression) is a potential method for ameliorating the diseased condition. In one embodiment of the disclosure, suppression of ABCA8 is achieved in a subject having a transporter malfunction due to overexpression of ABCA8.

Reduction of ABCA8 protein expression in a transgenic cell may be obtained for instance by introducing into cells an antisense construct based on the ABCA8 encoding sequence (SEQ ID NO: 1) or gene sequence or flanking regions thereof. In one specific, non-limiting embodiment, a nucleotide sequence from an *ABCA8* encoding sequence, *e.g.* SEQ ID NO: 1, is arranged in reverse orientation relative to the promoter sequence in the transformation vector. Other aspects of the vector may be chosen as discussed above (*see* Section VII).

The introduced sequence need not be the full-length human *ABCA8* cDNA (SEQ ID NO: 1) or gene, and need not be exactly homologous to the equivalent sequence found in the cell type to be transfected. Generally, however, where the introduced sequence is of shorter length, a higher degree of identity to the native ABCA8 sequence will be needed for effective antisense suppression. In other embodiments, the introduced antisense sequence in the vector may be at least 15 nucleotides in length, and improved antisense suppression typically will be observed as the length of the antisense sequence increases. In yet other embodiments, the length of the antisense sequence in the vector advantageously may be greater than 100 nucleotides, and can be up to about the full length of the human *ABCA8* cDNA or gene. In another embodiment, for suppression of the *ABCA8* gene itself, transcription of an antisense construct results in the production of RNA molecules that are the reverse complement of mRNA molecules transcribed from the endogenous *ABCA8* gene in the cell.

Although the exact mechanism by which antisense RNA molecules interfere with gene expression has not been elucidated, it is believed that antisense RNA molecules bind to the endogenous mRNA molecules and thereby inhibit translation of the endogenous mRNA.

Expression of ABCA8 can also be reduced using small inhibitory RNAs (siRNAs), for instance using techniques similar to those described previously (see, *e.g.*, Tuschl *et al.*, *Genes Dev* 13, 3191-3197, 1999; Caplen *et al.*, *Proc. Nat.l Acad. Sci. U. S. A.* 98, 9742-9747, 2001; and Elbashir *et al.*, *Nature* 411, 494-498, 2001).

In another embodiment, suppression of endogenous ABCA8 expression can be achieved using ribozymes. Ribozymes are synthetic RNA molecules that possess highly specific endoribonuclease activity. The production and use of ribozymes are disclosed in U.S. Patent No. 4,987,071 to Cech and U.S. Patent No. 5,543,508 to Haselhoff. In one embodiment, the inclusion of

ribozyme sequences within antisense RNAs may be used to confer RNA cleaving activity on the antisense RNA, such that endogenous mRNA molecules that bind to the antisense RNA are cleaved, which in turn leads to an enhanced antisense inhibition of endogenous gene expression.

In yet another embodiment, dominant negative mutant forms of ABCA8 may be used to block endogenous ABCA8 activity.

IX. Production of an Antibody to ABCA8 Protein

Monoclonal or polyclonal antibodies may be produced to the native ABCA8 protein or variant forms of this protein. In one embodiment, antibodies raised against an ABCA8 protein would specifically detect the ABCA8 protein. That is, such antibodies would recognize and bind the ABCA8 protein, or fragments thereof, and would not substantially recognize or bind to other proteins found in human cells. In other embodiments, antibodies against the human ABCA8 protein may recognize ABCA8 from other species (*e.g.*, murine ABCA8), and vice versa.

Monoclonal or polyclonal antibodies to the protein can be prepared as follows:

A. *Monoclonal Antibody Production by Hybridoma Fusion*

Monoclonal antibodies to epitopes of ABCA8 proteins can be prepared from murine hybridomas according to the classical method of Kohler and Milstein (*Nature* 256:495-497, 1975) or derivative methods thereof. In one specific, non-limiting embodiment, a mouse is repetitively inoculated with a few micrograms of the selected protein over a period of a few weeks. The mouse is then sacrificed, and the antibody-producing cells of the spleen isolated. The spleen cells are fused with mouse myeloma cells using polyethylene glycol, and the excess, non-fused, cells destroyed by growth of the system on selective media comprising aminopterin (HAT media). Successfully fused cells are diluted and aliquots of the dilution placed in wells of a microtiter plate, where growth of the culture is continued. Antibody-producing clones are identified by detection of antibody in the supernatant fluid of the wells by immunoassay procedures, such as ELISA, as originally described by Engvall (*Enzymol.* 70(A): 419-439, 1980), and derivative methods thereof. Selected positive clones can be expanded and their monoclonal antibody product harvested for use. Detailed procedures for monoclonal antibody production are described in Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988).

B. *Polyclonal Antibody Production by Immunization*

Polyclonal antiserum containing antibodies to heterogeneous epitopes of a single protein can be prepared by immunizing suitable animals with the expressed protein (for instance, expressed using a method described herein), which, in one specific, non-limiting embodiment, can be modified to enhance immunogenicity. Effective polyclonal antibody production is affected by many factors related both to the antigen and the host species. In one embodiment, small molecules may tend to be less immunogenic than others and may require the use of carriers and adjuvant, examples of which are known. In another embodiment, host animals may vary in response to site of inoculations and dose,

with either inadequate or excessive doses of antigen resulting in low titer antisera. In one specific, non-limiting embodiment, a series of small doses (ng level) of antigen administered at multiple intradermal sites may be most reliable. An effective immunization protocol for rabbits can be found in Vaitukaitis *et al. (J. Clin. Endocrinol. Metab.* 33: 988-991, 1971).

In one embodiment, booster injections will be given at regular intervals, and antiserum harvested when antibody titer thereof begins to fall, as determined semi-quantitatively (for example, by double immunodiffusion in agar against known concentrations of the antigen). See, for example, Ouchterlony *et al.* (In *Handbook of Experimental Immunology*, Wier, D. (ed.) chapter 19, Blackwell, 1973). In one specific, non-limiting embodiment the plateau concentration of antibody is usually in the range of about 0.1 to 0.2 mg/ml of serum (about 12 μM). Affinity of the antisera for the antigen is determined by preparing competitive binding curves, as described, for example, by Fisher (*Manual of Clinical Immunology*, Ch. 42, 1980).

C. ***Antibodies Raised against Synthetic Peptides***

A third approach to raising antibodies against ABCA8 proteins is to use synthetic peptides synthesized on a commercially available peptide synthesizer based upon the predicted amino acid sequence of the ABCA8 protein. Polyclonal antibodies can be generated by injecting such peptides into, for instance, rabbits.

D. ***Antibodies Raised by Injection of ABCA8 Encoding Sequence***

In one embodiment, antibodies may be raised against an ABCA8 protein by subcutaneous injection of a recombinant DNA vector that expresses the ABCA8 protein into laboratory animals, such as mice. In one specific, non-limiting embodiment, delivery of the recombinant vector into the animals may be achieved using a hand-held form of the Biolistic system (Sanford *et al., Particulate Sci. Technol.* 5:27-37, 1987), as described by Tang *et al.* (*Nature* 356:152-154, 1992). In other embodiments, expression vectors suitable for this purpose may include those that express the ABCA8 encoding sequence under the transcriptional control of either the human βactin promoter or the cytomegalovirus (CMV) promoter.

Antibody preparations prepared according to these protocols are useful in quantitative immunoassays which determine concentrations of antigen-bearing substances in biological samples; they are also used semi-quantitatively or qualitatively to identify the presence of antigen in a biological sample.

X. ABCA8 as a Transporter Protein

ABCA8 and its functional variants are believed to participate in the regulation of energy-dependent transport of a wide variety of substrates, including ions, metals, amino acids, lipids, peptides, sugars, vitamins, and steroid hormones across membranes. In one embodiment, the lack of expression of ABCA8 in a cell leads to a defect in transporter activity. In one specific, non-limiting embodiment, the administration of an ABCA8 recombinant nucleotide to a subject suffering from a transporter deficiency ameliorates the effects of the deficiency.

The expression of mutant transporter proteins can yield important information about the importance of each amino acid in the protein as well as the details of the mechanism of action of these proteins. Cells may have ABCA8 null mutations, ABCA8 missense mutations, or inactivation of ABCA8. In one embodiment, a mutant ABCA8 is expressed in a cell but is incapable of localizing to the correct subcellular location. In another embodiment, a mutant ABCA8 is incapable of binding to its intracellular binding partners. In yet another embodiment, a mutation in the upstream regulatory region of the ABCA8 gene abrogates the expression of the protein. In another embodiment, a mutant ABCA8 is incapable of transporting target substances.

It is possible that mutations in the ABCA8 gene that may lead to altered transporter function are not included in the cDNA but rather are located in other regions of the ABCA8 gene. Mutations located outside of the ORF that encode ABCA8 are not likely to affect the functional activity of the protein, but rather are likely to result in altered levels of the protein in cells. For example, mutations in the promoter region of the ABCA8 gene may prevent transcription of the gene and therefore lead to the complete absence of the ABCA8 protein, or absence of certain transcripts of the protein, in the cell.

Additionally, mutations within introns in the genomic sequence may also prevent expression of the ABCA8 protein or lead to expression of a non-functional form of the protein. Following transcription of a gene containing introns, the intron sequences are removed from the RNA molecule, in a process termed splicing, prior to translation of the RNA molecule that results in production of the encoded protein. When the RNA molecule is spliced to remove the introns, the cellular enzymes that perform the splicing function recognize sequences around the intron/exon border and in this manner recognize the appropriate splice sites. If a mutation exists within the sequence of the intron near and exon/intron junction, the enzymes may not recognize the junction and may fail to remove the intron. If this occurs, the encoded protein will likely be defective. Thus, mutations inside the intron sequences within the *ABCA8* gene (termed "splice site mutations") may also lead to defects in transporter activity. However, knowledge of the exon structure and intronic splice site sequences of the *ABCA8* gene is required to define the molecular basis of these abnormalities. The provision herein of the *ABCA8* cDNA sequence enables the cloning of the entire *ABCA8* gene (including the promoter and other regulatory regions and the intron sequences) and the determination of its nucleotide sequence. With this information in hand, diagnosis of a genetic predisposition to transporter deficiency based on DNA analysis will comprehend all possible mutagenic events at the *ABCA8* locus.

Compounds that modulate the expression or activity of ABCA8 can be used to regulate transporter activity. For instance, in some cases it may be determined that ABCA8 is expressed at low levels in a subject suffering from a transporter deficiency that arose as the result of the inefficient expression of ABCA8. Administration, to the subject, of an agent that up-regulates ABCA8 expression can improve transporter activity or ameliorate the transporter deficiency.

Changes in transporter activity can be assessed using a variety of *in vitro* and *in vivo* assays. These assays can be used to study the effect of ABCA8 on specific cell types or the effect of

particular mutations on ABCA8 transporter activity. By systematically introducing mutant ABCA8 constructs into cells and assessing their ability to transport substances across membranes, the importance of each amino acid for the protein's transporter activity can be determined. In addition, these assays can be used to screen for modulators of ABCA8 activity. The modulators identified in this way can then be used to alter ABCA8 expression in cells *in vitro* or *in vivo*.

In one embodiment, cells expressing wild-type ABCA8 are assessed for their ability to transport substances *in vitro*, as compared to cells that express a mutant ABCA8. In this embodiment, an *ABCA8* encoding nucleic acid construct is transfected into cells, and levels of the transported substance in these cells can be assayed and compared between wild-type and transfected cells for potential differences. Methods of analysis include, for instance, immobilization of the protein on columns to search for compounds that bind the immobilized protein (*see* Wainer *et al., J. Chromatogr. B. Biomed. Sci. Appl.* 724(1): 65-72, 1999) and disruption of the *ABCA8* gene by antisense or RNAi technology (*see* Paddison *et al., Proc. Natl. Acad. Sci. U. S. A.* 99(3): 1443-1448, 2002).

In yet another embodiment, samples or assays that are treated with a test compound that potentially modulates ABCA8 are compared to control samples that are not treated with the test compound, to examine the extent of modulation. In one embodiment, the compounds to be tested are present in the range from 0.1 nM to 10 mM. Control samples (untreated with modulators) are assigned a relative ABCA8 activity value of 100%. In one embodiment, inhibition of ABCA8 is achieved when the ABCA8 activity value relative to the control is about 90%. In other embodiments, inhibition of ABCA8 is achieved when the ABCA8 activity value relative to the control is about 75%, about 50%, about 25%, or about 5%. In another embodiment, activation of ABCA8 is achieved when the ABCA8 activity value relative to the control is about 110% (*e.g.* 10% more than the control). In other embodiments, activation of ABCA8 is achieved when the ABCA8 activity value relative to the control is about 150%, about 175% or about 200%.

The effect of test compounds upon ABCA8 activity can be assessed using the assays described above. Such assays include, but are not limited to, the ability to transport substances across membranes. In one embodiment, the compounds tested as modulators of ABCA8 are any small chemical compound, or biological entity, such as a polypeptide, sugar, nucleic acid or lipid. Furthermore, the ability of cells transfected with ABCA8 to transport fluorescent compounds can be assessed (*see* Gribar *et al., J. Membr. Biol.* 173(3): 203-214, 2000).

In another embodiment, the modulator is a genetically altered version of ABCA8. In other embodiments, the effect of potential modulators on ABCA8 protein or mRNA levels, transcriptional activation or repression of a reporter gene is measured.

XI. Nucleic Acid DNA-Based Diagnosis

The *ABCA8* sequence information presented herein can be used in the area of genetic testing for predisposition to transporter deficiency owing to defects in *ABCA8*, such as deletion, duplication or mutation of the *ABCA8* gene, or a portion thereof. The gene sequence of the *ABCA8* gene,

including intron-exon boundaries is also useful in such diagnostic methods. Whether an individual is carrying mutations in the *ABCA8* gene (or a portion thereof), or has a duplication(s) or heterozygous or homozygous deletion(s) of the *ABCA8* gene, may be detected at the DNA level with the use of a variety of techniques. For such a diagnostic procedure, a biological sample of the subject, which biological sample contains either DNA or RNA derived from the subject, is assayed for a mutated, duplicated or deleted *ABCA8* gene. Suitable biological samples include samples containing genomic DNA or RNA obtained from body cells, such as those present in peripheral blood, urine, saliva, tissue biopsy, surgical specimen, amniocentesis samples and autopsy material. Biological samples can be obtained from normal, healthy subjects or from subjects who are predisposed to or who are suffering from any one of a variety of the effects of transporter deficiencies such as, but not limited to, hypercholesterolemia, or defective transport of hydrophobic compounds such as lipids, sterols or fatty acids.

The detection in the biological sample of either a mutant *ABCA8* gene, a mutant *ABCA8* RNA, or a duplicated or homozygously or heterozygously deleted *ABCA8* gene, may be performed by a number of methodologies, examples of which are discussed below.

One embodiment of such detection techniques for the identification of unknown mutations is the amplification (*e.g.*, polymerase chain reaction amplification) of reverse transcribed RNA (RT-PCR) isolated from a subject, followed by direct DNA sequence determination of the products. The presence of one or more nucleotide differences between the obtained sequence and the prototypical *ABCA8* cDNA sequence, and especially, differences in the open reading frame (ORF) portion of the nucleotide sequence, are taken as indicative of a potential *ABCA8* gene mutation.

By way of example, the disclosure includes a method of amplifying a nucleic acid and, more particularly, a nucleic acid including any one of SEQ ID NO: 1, a fragment containing residues 1927-2046 of SEQ ID NO: 1, and/or of a complementary nucleotide sequence thereof or a variant thereof, contained in a sample, the method involving the steps of:

a) bringing the sample in which the presence of a target nucleic acid is suspected into contact with a pair of nucleotide primers whose hybridization position is located respectively on the 5' side and on the 3' side of the region of the target nucleic acid whose amplification is sought, in the presence of the reagents necessary for the amplification reaction; and

b) detecting the amplified nucleic acids (if any).

The nucleotide primers according to the disclosure are useful in methods of genotyping subjects and/or of genotyping populations, in particular in the context of studies of association between particular allele forms or particular forms of groups of alleles (haplotypes) in subjects and the existence of a particular phenotype (character) in these subjects. For example, primers can be used to detect the predisposition of these subjects to develop diseases linked to a deficiency of cholesterol reverse transport and inflammation signaling lipids, or alternatively the predisposition of these subjects to develop a pathology whose candidate chromosomal region is situated on chromosome 17, for example on the 17q arm, and such as in the 17q24 locus.

Alternatively, DNA extracted from a biological sample may be used directly for amplification. Direct amplification from genomic DNA would be appropriate for analysis of the entire *ABCA8* gene including regulatory sequences located upstream and downstream from the open reading frame, or intron/exon borders. Reviews of direct DNA diagnosis have been presented by Caskey (*Science* 236:1223-1228, 1989) and by Landegren *et al.* (*Science* 242:229-237, 1989).

Other mutation scanning techniques appropriate for detecting unknown mutations within amplicons derived from DNA or cDNA could also be performed. These techniques include direct sequencing (without sequencing), single-strand conformational polymorphism analysis (SSCP) (for instance, see Hongyo *et al.*, *Nucleic Acids Res.* 21:3637-3642, 1993), chemical cleavage (including HOT cleavage) (Bateman *et al.*, *Am. J. Med. Genet.* 45:233-240, 1993; reviewed in Ellis *et al.*, *Hum. Mutat.* 11:345-353, 1998), denaturing gradient gel electrophoresis (DGGE), ligation amplification mismatch protection (LAMP), and enzymatic mutation scanning (Taylor and Deeble, *Genet. Anal.* 14:181-186, 1999), followed by direct sequencing of amplicons with putative sequence variations.

If studies of *ABCA8* genes/coding sequences isolated from biological samples reveal particular mutations, genomic amplifications, or deletions, which occur at a high frequency within a population of individuals, DNA diagnostic methods can be designed to specifically detect the most common, or most closely disease-linked, ABCA8 defects.

The detection of specific DNA mutations may be achieved by methods such as hybridization using allele specific oligonucleotides (ASOs) (Wallace *et al.*, *CSHL Symp. Quant. Biol.* 51:257-261, 1986), direct DNA sequencing (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* 81:1991-1995, 1988), the use of restriction enzymes (Flavell *et al.*, *Cell* 15:25-41, 1978; Geever *et al.*, 1981), discrimination on the basis of electrophoretic mobility in gels with denaturing reagent (Myers and Maniatis, Cold Spring Harbor Symp. Quant. Biol. 51:275-284, 1986), RNase protection (Myers *et al.*, *Science* 230:1242-1246, 1985), chemical cleavage (Cotton *et al.*, *Proc. Natl. Acad. Sci. USA* 85:4397-4401, 1985), and the ligase-mediated detection procedure (Landegren *et al.*, *Science* 241:1077-1080, 1988).

Oligonucleotides specific to normal or mutant sequences are chemically synthesized using commercially available machines. These oligonucleotides are optionally labeled radioactively with isotopes (such as ^{32}P) or non-radioactively, with tags such as biotin (Ward and Langer, *Proc. Natl. Acad. Sci. USA* 78:6633-6657, 1981), and hybridized to individual DNA samples immobilized on membranes or other solid supports by dot-blot or transfer from gels after electrophoresis. These specific sequences are visualized by methods such as autoradiography or fluorometric (Landegren *et al.*, *Science* 242:229-237, 1989) or colorimetric reactions (Gebeyehu *et al.*, *Nucleic Acids Res.* 15:4513-4534, 1987). Using an antisense oligonucleotide (ASO) specific for a normal allele, the absence of hybridization would indicate a mutation in the particular region of the gene, or deleted *ABCA8* gene. In contrast, if an ASO specific for a mutant allele hybridizes to a clinical sample, this would indicate the presence of a mutation in the region defined by the ASO.

Sequence differences between normal and mutant forms of the ABCA8 gene may also be

revealed by the direct DNA sequencing method of Church and Gilbert (*Proc. Natl. Acad. Sci. USA* 81:1991-1995, 1988). Cloned DNA segments may be used as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined *in vitro* with nucleic acid amplification, *e.g.*, PCR (Wrichnik *et al., Nucleic Acids Res.* 15:529-542, 1987; Wong *et al., Nature* 330:384-386, 1987; Stoflet *et al., Science* 239:491-494, 1988). In this approach, a sequencing primer that lies within the amplified sequence is used with double-stranded PCR product or single-stranded template generated by a modified PCR. The sequence determination is performed by conventional procedures with radiolabeled nucleotides or by automatic sequencing procedures with fluorescent tags.

Nucleotide probes and primers hybridizing with an ABCA8 nucleic acid (genomic DNA, messenger RNA, cDNA) are encompassed by the disclosure, and are suitable for use with the methods discussed above. A nucleotide primer or probe may be prepared by any suitable method well known to persons skilled in the art, including by cloning and action of restriction enzymes or by direct chemical synthesis according to techniques such as the phosphodiester method by Narang *et al.* (1979, *Methods Enzymol*, 68:90-98) or by Brown *et al.* (1979, *Methods Enzymol*, 68:109-151), the diethylphosphoramidite method by Beaucage *et al.* (1981, *Tetrahedron Lett*, 22: 1859-1862) or the technique on a solid support described in EU patent No. EP 0,707,592.

Each of the nucleic acids according to the disclosure, including oligonucleotide probes and primers, may be labeled, if desired, by incorporating a marker that can be detected by spectroscopic, photochemical, biochemical, immunochemical or chemical means. For example, such markers may consist of radioactive isotopes (such as ^{32}P, ^{33}P, ^{3}H, ^{35}S), fluorescent molecules (such as 5-bromodeoxyuridine, fluorescein, acetylaminofluorene, digoxigenin) or ligands such as biotin. The labeling of the probes can be carried out for instance by incorporating labeled molecules into the polynucleotides by primer extension, or alternatively by addition to the 5' or 3' ends. Examples of nonradioactive labeling of nucleic acid fragments are described in particular in French patent No. 78 109 75 or in the articles by Urdea *et al.* (*Nucleic Acids Research*, 11:4937-4957, 1988) or Sanchez-Pescador *et al.* (*J. Clin. Microbiol.*, 26(10):1934-1938, 1988).

Nucleotide probes and primers according to embodiments of the disclosure may have structural characteristics of the type to allow amplification of the signal, such as the probes described by Urdea *et al.* (*Nucleic Acids Symp Ser.*, 24:197-200, 1991) or alternatively in European patent No. EP-0,225,807.

Oligonucleotides according to the disclosure may be used in particular in Southern-type hybridizations with the genomic DNA or alternatively in northern-type hybridizations with the corresponding messenger RNA when the expression of the corresponding transcript is sought in a sample, as described herein. The probes and primers according to the disclosure may also be used for the detection of products of *in vitro* amplification or alternatively for the detection of mismatches.

Nucleotide probes or primers according to the disclosure may be immobilized on a solid support. Such solid supports are well known to persons skilled in the art and comprise surfaces of

wells of microtiter plates, polystyrene beads, magnetic beads, nitrocellulose bands or microparticles such as latex particles as well as glass or polycarbonate slides or various substances. The probes according to the disclosure, immobilized on a support, may be ordered into matrices such as "DNA chips." Such ordered matrices have in particular been described in US patent No. 5,143,854, and in published PCT applications WO 90/15070 and WO 92/10092. Support matrices on which oligonucleotide probes have been immobilized at a high density are for example described in US Patent No. 5,412,087 and in published PCT application WO 95/11995.

According to the disclosure, nucleic acid fragments derived from a polynucleotide including SEQ ID NO: 1 or a complementary nucleotide sequence, or including residues 1927-2046 of SEQ ID NO: 1, are useful as probes for the detection of the presence of at least one copy of a nucleotide sequence of the *ABCA8* gene or of a fragment or of a variant (*e.g.*, containing a mutation or a polymorphism) thereof in a sample. The nucleotide probes or primers according to the disclosure comprise a contiguous nucleotide sequence found in SEQ ID NO: 1, residues 1927-2046 of SEQ ID NO: 1, or a complementary nucleotide thereof. Particular embodiments include nucleotide probes or primers that include at least 8 consecutive nucleotides having a sequence as shown in residues 1927-2046 of SEQ ID NO: 1 or a complementary nucleotide sequence thereof. Certain provided nucleotide probes or primers have a length of 10, 12, 15, 18 or 20, to 25, 35, 40, 50, 70, 80, 100 consecutive nucleotides or more of a nucleic acid, which comprise a particular sequence as found in residues 1927-2046 of SEQ ID NO: 1 or a complementary nucleotide sequence thereof. Nucleotide probes or primers also include oligonucleotides that hybridize, under the high stringency hybridization conditions as discussed herein, with a nucleic acid composing SEQ ID NO: 1, residues 1927-2046 of SEQ ID NO: 1, or a complementary nucleotide sequence thereof.

The present disclosure also relates to a method of detecting the presence of a nucleic acid including a nucleotide sequence of any one of SEQ ID NO: 1, a fragment containing residues 1927-2046 of SEQ ID NO: 1, and/or of a complementary nucleotide sequence thereof, or a nucleic acid fragment or variant of any one of SEQ ID NO: 1, a fragment containing residues 1927-2046 of SEQ ID NO: 1, and/or of a complementary nucleotide sequence thereof in a sample, which method involves: 1) bringing one or more nucleotide probes or primers into contact with the sample to be tested; and 2) detecting the complex that may have formed between the probe(s) and the nucleic acid present in the sample.

In some embodiments, the oligonucleotide probes and/or primers are immobilized on a support. In some embodiments, the oligonucleotide probes and/or primers include a detectable marker.

The nucleotide primers according to the disclosure may be used to amplify any one of the nucleic acids according to the disclosure, and more particularly a nucleic acid including a nucleotide sequence of any one of SEQ ID NO: 1, a fragment containing residues 1927-2046 of SEQ ID NO: 1, and/or of a complementary nucleotide sequence thereof. Alternatively, the nucleotide primers according to the disclosure may be used to amplify a nucleic acid fragment or variant of any one of

SEQ ID NO: 1, a fragment containing residues 1927-2046 of SEQ ID NO: 1, and/or of a complementary nucleotide sequence thereof. Amplification in this manner may be used to detect a target nucleic acid of interest or alternatively to detect mutations in the coding regions or the non-coding regions of the nucleic acids according to the disclosure.

Sequence alterations may occasionally generate fortuitous restriction enzyme recognition sites or may eliminate existing restriction sites. Changes in restriction sites are revealed by the use of appropriate enzyme digestion followed by conventional gel-blot hybridization (Southern, *J. Mol. Biol.* 98:503-517, 1975). DNA fragments carrying the restriction site (either normal or mutant) are detected by their reduction in size or increase in corresponding restriction fragment numbers. Genomic DNA samples may also be amplified by PCR prior to treatment with the appropriate restriction enzyme; fragments of different sizes are then visualized under UV light in the presence of ethidium bromide after gel electrophoresis.

Genetic testing based on DNA sequence differences may be achieved by detection of alteration in electrophoretic mobility of DNA fragments in gels, with or without denaturing reagent. Small sequence deletions and insertions can be visualized by high-resolution gel electrophoresis. For example, a PCR product with small deletions is clearly distinguishable from a normal sequence on an 8 % non-denaturing polyacrylamide gel (WO 91/10734; Nagamine *et al., Am. J. Hum. Genet.* 45:337-339, 1989). DNA fragments of different sequence compositions may be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific "partial-melting" temperatures (Myers *et al., Science* 230:1242-1246, 1985). Alternatively, a method of detecting a mutation comprising a single base substitution or other small change could be based on differential primer length in a PCR. For example, an invariant primer could be used in addition to a primer specific for a mutation. The PCR products of the normal and mutant genes can then be differentially detected in acrylamide gels. Another method, single-strand conformation polymorphism (SSCP), is based on the fact that a single-base substitution alters the conformation of single-stranded DNA under non-denaturing conditions. Altered conformation affects the migration velocity of single-stranded DNA, which is detected as shifted or new bands on a non-denaturing gel. The mutations underlying the shifted or new bands are then characterized by sequencing.

In addition to conventional gel-electrophoresis and blot-hybridization methods, DNA fragments may also be visualized by methods in which the individual DNA samples are not immobilized on membranes. The probe and target sequences may be both in solution, or the probe sequence may be immobilized (Saiki *et al., Proc. Nat. Acad. Sci. USA* 86: 6230-6234, 1989). A variety of detection methods, such as autoradiography involving radioisotopes, direct detection of radioactive decay (in the presence or absence of scintillant), spectrophotometry involving calorigenic reactions and fluorometry involved fluorogenic reactions, may be used to identify specific individual genotypes.

If multiple mutations are encountered frequently in the *ABCA8* gene, a system capable of

detecting such multiple mutations likely will be desirable. For example, a nucleic acid amplification reaction with multiple, specific oligonucleotide primers and hybridization probes may be used to identify all possible mutations at the same time (Chamberlain *et al., Nucl. Acids Res.* 16:1141-1155, 1988). The procedure may involve immobilized sequence-specific oligonucleotide probes (Saiki *et al., Proc. Nat. Acad. Sci. USA* 86:6230-6234, 1989).

Expression levels of the *ABCA8* gene can also be determined by methods such as Northern or Southern blot analysis using labeled oligonucleotides specific to normal or mutant sequences. These oligonucleotides are labeled radioactively with isotopes (such as ^{32}P) or non-radioactively, with tags such as biotin (Ward and Langer, *Proc. Natl. Acad. Sci. USA* 78: 6633-6657, 1981), and hybridized to individual DNA samples immobilized on membranes or other solid supports by dot-blot or transfer from gels after electrophoresis. Quantitative or semi-quantitative PCR can also be used to measure the amount of ABCA8 cDNA in a sample using ABCA8 oligonucleotide primers. Visualization methods such as autoradiography or fluorometric (Landegren *et al., Science* 242: 229-237, 1989) or colorimetric reactions (Gebeyehu *et al., Nucleic Acids Res.* 15: 4513-4534, 1987) can be used to detect a signal and the signals quantitated using, for instance, a spectrophotometer, a scintillation counter, a densitometer or a Phosphorimager (Amersham Biosciences). The Phosphorimager is able to analyze both DNA and protein samples from blots and gels using autoradiographic, direct fluorescence or chemifluorescence detection. Because the Phosphorimager is more sensitive than ordinary x-ray film, exposure times can be reduced up to ten-fold and signal quantitation of both weak and strong signals on the same blot is possible. Images can be visualized and evaluated with the aid of computer programs such as ImageQuant™.

XII. Qualitative and Quantitative Detection of ABCA8 Protein

The *ABCA8* sequence information presented herein is useful in detecting the presence or absence of ABCA8 in cultured cells or primary cells. Quantitative and qualitative methods of detection of proteins are well-known in the art, and are discussed herein. Quantitative detection is useful for diagnosing over- or underexpression of ABCA8 proteins in a subject, while qualitative information gives, for example, information regarding tissue types in which ABCA8 may be expressed. For such qualitative or quantitative assessment, a biological sample of the subject, which biological sample contains either DNA or RNA derived from the subject, is assayed for the presence or absence of ABCA8. Suitable biological samples include samples containing genomic DNA or RNA obtained from body cells, such as those present in peripheral blood, urine, saliva, tissue biopsy, surgical specimen, amniocentesis samples and autopsy material. Biological samples can be obtained from normal, healthy subjects or from subjects who are predisposed to or who are suffering from any one of a variety of the effects of transporter deficiencies such as, but not limited to, hypercholesterolemia, or defective transport of hydrophobic compounds such as lipids, sterols or fatty acids.

Antibodies can be used to assess the presence or absence of ABCA8 proteins in cultured

cells or primary cells. The determination whether an antibody specifically detects ABCA8 proteins is made by any one of a number of standard immunoassay methods; for instance, the Western blotting technique (Sambrook *et al.*, In *Molecular Cloning: A Laboratory Manual*, CSHL, New York, 1989).

In one embodiment, it is determined whether a given antibody preparation (such as one produced in a mouse) specifically detects ABCA8 proteins by Western blotting. In one specific, non-limiting embodiment total cellular protein is extracted from human cells (for example, lymphocytes) and electrophoresed on a sodium dodecyl sulfate-polyacrylamide gel.

In another embodiment, the cellular protein is extracted from a biological sample. The proteins are then transferred to a membrane (for example, nitrocellulose or PVDF) by Western blotting, and the antibody preparation is incubated with the membrane. After washing the membrane to remove non-specifically bound antibodies, the presence of specifically bound antibodies is detected by the use of (by way of example) an anti-mouse antibody conjugated to an enzyme such as alkaline phosphatase. Application of an alkaline phosphatase substrate 5-bromo-4-chloro-3-indolyl phosphate/nitro blue tetrazolium results in the production of a dense blue compound by immunolocalized alkaline phosphatase. Antibodies that specifically detect ABCA8 proteins will, by this technique, be shown to bind to ABCA8 protein bands (which will be localized at a given position on the gel determined by its molecular weight). Non-specific binding of the antibody to other proteins may occur and may be detectable as a weak signal on the Western blot. The non-specific nature of this binding will be recognized by one skilled in the art by the weak signal obtained on the Western blot relative to the strong primary signal arising from the specific antibody-ABCA8 protein binding.

In one embodiment, substantially pure ABCA8 protein suitable for use as an immunogen is isolated from the transfected cells as described above. In one specific, non-limiting embodiment the concentration of protein in the final preparation is adjusted, for example, by concentration on an Amicon (Millipore, Bedford, Massachusetts) or similar filter device, to the level of a few micrograms per milliliter.

In other embodiments, antibodies against ABCA8 proteins are used to localize ABCA8 to specific cell types or to specific subcellular locations in immunohistochemical or immunofluorescence assays. In one embodiment, the cells are selected from a variety of cell lines. In other embodiments, primary cells are isolated from a subject and are maintained in culture or the sample is sectioned and the sections are prepared directly for immunohistochemistry or immunofluorescence. In one specific, non-limiting embodiment, the cells are fixed, incubated in a blocking medium, incubated with the antibody directed against ABCA8 followed by a second incubation with a secondary antibody that is conjugated to a fluorescent probe or a colorimetric agent. Cells that express an ABCA8 protein that is recognized by the antibody exhibit a color or are fluorescent when viewed under a light or fluorescence microscope, respectively.

An alternative method of diagnosing ABCA8 gene deletion, amplification, or mutation is to quantitate the level of ABCA8 protein in the cells of a subject. In one embodiment, this diagnostic

tool would be useful for detecting reduced levels of ABCA8 protein that result from, for example, mutations in the promoter regions of the ABCA8 gene or mutations within the coding region of the gene that produce truncated, non-functional or unstable polypeptides, as well as from deletions of the entire ABCA8 gene. In another embodiment, duplications of the ABCA8 gene may be detected as an increase in the expression level of this protein. The determination of reduced or increased ABCA8 protein levels would be an alternative or supplemental approach to the direct determination of ABCA8 gene deletion, duplication or mutation status by the methods outlined above.

The availability of antibodies specific to ABCA8 proteins will facilitate the quantitation of cellular ABCA8 proteins by one of a number of immunoassay methods, which are well known in the art and are presented herein and in, for instance, Harlow and Lane (*Antibodies, A Laboratory Manual*, CSHL, New York, 1988). Many techniques are commonly known in the art for the detection and quantification of antigen. In one specific, non-limiting embodiment, the purified antigen will be bound to a substrate, the antibody of the sample will bind via its Fab portion to this antigen, the substrate will then be washed and a second, labeled antibody will then be added, which will bind to the Fc portion of the antibody that is the subject of the assay. The second, labeled antibody will be species specific, *i.e.*, if the serum is from a rabbit, the second, labeled antibody will be anti-rabbit-IgG antibody. The specimen will then be washed and the amount of the second, labeled antibody that has been bound will be detected and quantified by standard methods.

Examples of methods for the detection of antibodies in biological samples, including methods employing dip strips or other immobilized assay devices, are disclosed for instance in the following patents: U.S. Patents No. 5,965,356 (Herpes simplex virus type specific seroassay); 6,114,179 (Method and test kit for detection of antigens and/or antibodies); 6,077,681 (Diagnosis of motor neuropathy by detection of antibodies); 6,057,097 (Marker for pathologies comprising an auto-immune reaction and/or for inflammatory diseases); and 5,552,285 (Immunoassay methods, compositions and kits for antibodies to oxidized DNA bases).

For the purposes of quantitating the ABCA8 protein, a biological sample of the subject, which sample includes cellular proteins, can be used. Such a biological sample may be obtained from body cells, such as those present in peripheral blood, urine, saliva, tissue biopsy, amniocentesis samples, surgical specimens and autopsy material. Biological samples can be obtained from normal, healthy subjects or from subjects who are predisposed to or who are already suffering from any one of a variety of transporter deficiencies, such as, but not limited to, defective transport of cholesterol, fatty acids, or lipids, or the effects of defective steroid hormones derived from the cholesterol of individuals with mutant ABCA8.

Quantitation of ABCA8 protein can be achieved for instance by immunoassay and compared to levels of the protein found in healthy cells (*e.g.*, cells from a subject known not to suffer from a transporter deficiency). In one embodiment, a significant (*e.g.*, 10% or greater, for instance, 20%, 25%, 30%, 50% or more) reduction in the amount of ABCA8 protein in the cells of a subject compared to the amount of ABCA8 protein found in normal human cells would be taken as an

indication that the subject may have deletions or mutations in the ABCA8 gene locus, whereas in another embodiment, a significant (*e.g.*, 10% or greater, for instance, 20%, 25%, 30%, 50% or more) increase would indicate that a duplication or enhancing mutation had occurred.

XIII. ABCA8 *Knockout and Overexpression Transgenic Animals*

Mutant organisms that under-express or over-express ABCA8 proteins are useful for research. Such mutants allow insight into the physiological and/or pathological role of ABCA8 in a healthy and/or pathological organism via study using a mutant model system (*e.g.*, a transgenic ABCA8 knockout mouse). These mutants are "genetically engineered," meaning that information in the form of nucleotides has been transferred into the mutant's genome at a location, or in a combination, in which it would not normally exist. Nucleotides transferred in this way are said to be "non-native." In one embodiment, a non-*ABCA8* promoter inserted upstream of a native *ABCA8* gene would be non-native. In other embodiments, an extra copy of an *ABCA8* gene or other encoding sequence on a plasmid, transfected into a cell, would be non-native, whether that extra copy was *ABCA8* derived from the same, or a different species.

Mutants may be, for example, produced from mammals, such as mice, that either over-express or under-express ABCA8 protein, or that do not express ABCA8 at all. In one embodiment, over-expression mutants are made by increasing the number of *ABCA8*-encoding sequences (such as genes) in the organism. In other embodiments, over-expression mutants are made by introducing an ABCA8-encoding sequence into the organism under the control of a constitutive or inducible or viral promoter such as the mouse mammary tumor virus (MMTV) promoter or the whey acidic protein (WAP) promoter or the metallothionein promoter. In yet other embodiments, mutants that under-express ABCA8 may be made by using an inducible or repressible promoter, or by deleting the *ABCA8* gene, or by destroying or limiting the function of the *ABCA8* gene, for instance by disrupting the gene by transposon insertion.

In another embodiment, antisense genes may be engineered into the organism, under a constitutive or inducible promoter, to decrease or prevent ABCA8 expression, as discussed above.

A gene is "functionally deleted" when genetic engineering has been used to negate or reduce gene expression to negligible levels. When a mutant is referred to in this application as having the *ABCA8* gene altered or functionally deleted, this refers to the *ABCA8* gene and to any ortholog of this gene. When a mutant is referred to as having "more than the normal copy number" of a gene, this means that it has more than the usual number of genes found in the wild-type organism, *e.g.*, in the diploid mouse or human.

In one embodiment, a mutant mouse over-expressing ABCA8 may be made by constructing a plasmid having the *ABCA8* gene driven by a promoter, such as the mouse mammary tumor virus (MMTV) promoter or the whey acidic protein (WAP) promoter. In one specific, non-limiting embodiment, this plasmid may be introduced into mouse oocytes by microinjection. The oocytes are implanted into pseudopregnant females, and the litters are assayed for insertion of the transgene.

Multiple strains containing the transgene are then available for study.

WAP is quite specific for mammary gland expression during lactation, and MMTV is expressed in a variety of tissues including mammary gland, salivary gland and lymphoid tissues. In other embodiments, other promoters might be used to achieve various patterns of expression, *e.g.*, the metallothionein promoter.

In another embodiment, an inducible system may be created in which the subject expression construct is driven by a promoter regulated by an agent that can be fed to the mouse, such as tetracycline. Such techniques are well known in the art.

In yet another embodiment, a mutant knockout animal (*e.g.*, mouse) from which the *ABCA8* gene is deleted or otherwise disabled can be made by removing coding regions of the *ABCA8* gene from embryonic stem cells. The methods of creating deletion mutations by using a targeting vector have been described (see, for instance, Thomas and Capecch, *Cell* 51: 503-512, 1987).

In one embodiment, knockout mice are used as hosts to test the effects of various ABCA8 constructs on cell growth. In other embodiments, transgenic mice with the endogenous *ABCA8* gene knocked-out can be used in an assay to screen for compounds that modulate the ABCA8 activity. A transgenic mouse that is heterozygous or homozygous for integrated transgenes that have functionally disrupted the endogenous *ABCA8* gene can be used as a sensitive *in vivo* screening assay for the ABCA8 ligands and modulators of ABCA8 activity.

XIV. Nucleic Acid-Based ABCA8 Therapy

Medical genetic approaches for combating ABCA8-mediated defects in subjects, such as a transporter deficiency, are now made possible.

In one embodiment, retroviruses are a preferred vector for experiments in medical genetics, as they yield a high efficiency of infection and stable integration and expression (Orkin *et al., Prog. Med. Genet.* 7:130-142, 1988). In one specific, non-limiting embodiment, the full-length *ABCA8* gene or cDNA can be cloned into a retroviral vector and driven from either its endogenous promoter or, for instance, from the retroviral LTR (long terminal repeat). In other embodiments, viral transfection systems may also be utilized for this type of approach, including adenovirus, adeno-associated virus (AAV) (McLaughlin *et al., J. Virol.* 62:1963-1973, 1988), Vaccinia virus (Moss *et al., Annu. Rev. Immunol.* 5:305-324, 1987), Bovine Papilloma virus (Rasmussen *et al., Methods Enzymol.* 139:642-654, 1987) or members of the herpesvirus group such as Epstein-Barr virus (Margolskee *et al., Mol. Cell. Biol.* 8:2837-2847, 1988).

Medical genetic techniques include the use of RNA-DNA hybrid oligonucleotides, as described by Cole-Strauss *et al.* (*Science* 273:1386-1389, 1996). This technique may allow for site-specific integration of cloned sequences, thereby permitting accurately targeted gene replacement.

In addition to delivery of *ABCA8* to cells using viral vectors, it is possible to use non-infectious methods of delivery. In one embodiment, lipidic and liposome-mediated gene delivery will be used for transfection of various genes (for reviews, see Templeton and Lasic, *Mol. Biotechnol.*

11:175-180, 1999; Lee and Huang, *Crit. Rev. Ther. Drug Carrier Syst.* 14:173-206; and Cooper, *Semin. Oncol.* 23:172-187, 1996). In another embodiment, cationic liposomes will be used as a viable alternative to the viral vectors (de Lima *et al.*, *Mol. Membr. Biol.* 16:103-109, 1999). In yet other embodiments, cationic liposomes can be targeted to specific cells through the inclusion of, for instance, monoclonal antibodies or other appropriate targeting ligands (Kao *et al.*, *Cancer Gene Ther.* 3:250-256, 1996).

XV. Pharmaceutical Compositions

Also disclosed herein is a pharmaceutical composition intended for the prevention of or treatment of a patient or a subject affected with ABCA8 deficiency, including a recombinant vector according to the disclosure, in combination with one or more physiologically compatible excipients. The disclosure also relates to the use of a nucleic acid according to the disclosure, encoding the ABCA8 protein or a variant or fragment thereof, for the manufacture of a medicament intended for the prevention of a disease, for instance atherosclerosis or arteriosclerosis in various forms or more particularly for the treatment of subjects affected by a dysfunction of cholesterol reverse transport or inflammatory lipophilic substances transport.

The disclosure also relates to the use of a recombinant vector according to the disclosure, including an *ABCA8* nucleic acid, for the manufacture of a medicament intended for the prevention of a disease, such as atherosclerosis or arteriosclerosis in various forms, or in some embodiments particularly for the treatment of subjects affected by a dysfunction of cholesterol reverse transport or inflammatory lipophilic substances transport.

The present disclosure also relates to a pharmaceutical composition including one or more defective recombinant viruses as described above.

Pharmaceutical compositions disclosed herein may be formulated for administration by the topical, oral, parenteral, intranasal, intravenous, intramuscular, subcutaneous, intraocular or transdermal route and so forth. In some embodiments, the pharmaceutical compositions comprise a pharmaceutically acceptable vehicle or physiologically compatible excipient for an injectable formulation, in particular for an intravenous injection, such as for example into the subject's portal vein. These may relate in particular to isotonic sterile solutions or dry, in particular, freeze-dried, compositions that, upon addition depending on the case of sterilized water or physiological saline, allow the preparation of injectable solutions.

XVI. Methods of Screening for Agonist or Antagonist Compounds

The disclosure further relates to various methods of screening compounds or small molecules for therapeutic use that are useful in the treatment of diseases caused by a deficiency in ABCA8, further in the transport of cholesterol, or more particularly, in inflammatory lipid substances.

The disclosure therefore also relates to the use of the ABCA8 polypeptide, or of cells expressing the ABCA8 polypeptide, for screening active ingredients for the prevention and/or

treatment of diseases resulting from a dysfunction in ABCA8 deficiency. Peptides from functional sites (for instance, from within ABC trans-membrane and nucleotide-binding domains as described herein) and oligopeptide or immunogenic fragments of ABCA8, for instance oligopeptides within the region encompassed by residues 597-636 of SEQ ID NO: 2, can serve for screening product libraries using existing techniques. The polypeptide fragment(s) used in this type of screening may be free in solution, bound to a solid support, at the cell surface, or in the cell. The formation of the binding complexes between the ABCA8 polypeptide fragment and the tested agent can then be measured.

Another product screening technique that may be used in high-flux screenings giving access to products having affinity for the protein of interest is described in published International Application WO 84/03564. In this method, applied to the ABCA8 protein, various products are synthesized on a solid surface. These products react with an ABCA8 protein or a fragment thereof, and the surface is washed. The products binding ABCA8 protein/fragment are then detected by methods known to persons skilled in the art. Antibodies can also be used to capture a peptide and immobilize it on a support.

Another embodiment is to perform a product screening method, using ABCA8 neutralizing competition antibodies in solution with the ABCA8 protein or polypeptide and a compound that potentially binds the ABCA8 protein. In this manner, the antibodies may be used to detect the presence of a peptide having a common antigenic unit with an ABCA8 polypeptide or protein.

Accordingly, this disclosure relates to the use of any method of screening products (*i.e.*, compounds, small molecules, and so forth), based on translocation of a substrate, such as cholesterol or a lipophilic substance, between membranes or vesicles. Such translocation can be of the synthetic or cellular type (*e.g.*, mammal cells, insect cells, bacteria, or yeasts expressing constitutively or having incorporated human ABCA8 encoding nucleic acids). To this effect, labeled lipophilic substance analogs may be used.

Furthermore, knowing that the disruption of numerous transporters has been described (van den Hazel *et al.*, *J. Biol. Chem.*, 274: 1934-41, 1999), it is contemplated that cellular mutants having a characteristic phenotype can also be used. The function of such cells can be complemented with the ABCA8 protein and the whole used for screening purposes.

The disclosure also relates to a method of screening a compound, or small molecule active or potentially active in the transport of lipophilic substances, for activity as an agonist or antagonist of an ABCA8 polypeptide, the method including the following:

a) preparing a membrane vesicle including an ABCA8 polypeptide(s) and a substrate, which substrate optionally includes a detectable marker;

b) incubating the vesicle with a candidate agonist or antagonist compound;

c) qualitatively and/or quantitatively measuring release of the substrate from the vesicles; and

d) comparing the release measurement with a measurement of release of labeled substrate by a vesicle that has not been previously incubated with the agonist or antagonist candidate compound.

A membrane vesicle can be a synthetic lipid vesicle, which may be prepared according to techniques well known to a person skilled in the art. In some embodiments, the ABCA8 protein is recombinant protein.

In some embodiments, the membrane vesicle is a vesicle of a plasma membrane derived from cells expressing the ABCA8 polypeptide. These may be cells naturally expressing the ABCA8 polypeptide or cells transfected with a nucleic acid encoding an ABCA8 polypeptide or recombinant vector including a nucleic acid encoding an ABCA8 polypeptide.

In some of the provided methods, the substrates are lipid substrates. Lipid substrates used in the above screening method may include (but are not limited to) prostaglandins or prostacyclins, cholesterol or phosphatidylcholine, radioactively labeled lipid substrates (*e.g.*, with an isotope chosen from ^{3}H or ^{125}I), and fluorescently labeled compounds (*e.g.*, labeled with NBD or pyrene). In labeled compounds, the measurement of the fluorescence or of the radioactivity released by the vesicle is a direct reflection of the activity of substrate transport by the ABCA8 polypeptide.

The membrane vesicle including the labeled substances and the ABCA8 polypeptide can be immobilized at the surface of a solid support prior to incubating them with the candidate compound.

The disclosure also relates to a method of screening a compound or small molecule active (or thought to potentially be active) on the transport of cholesterol or lipid substances, an agonist or antagonist of the ABCA8 polypeptide, the method including the following:

a) obtaining cells, for example a cell line that, either naturally or after transfecting the cell with a ABCA8 encoding nucleic acid, expresses the ABCA8 polypeptide;

b) incubating the cells of a) in the presence of an anion labeled with a detectable marker;

c) washing the cells of b) in order to remove the excess of the labeled anion which has not penetrated into these cells;

d) incubating the cells obtained in c) with an agonist or antagonist candidate compound for the ABCA8 polypeptide;

e) measuring efflux of the labeled anion; and

f) comparing the value of efflux of the labeled anion determined in e) with a value of the efflux of a labeled anion measured with cells that have not been previously incubated in the presence of the agonist or antagonist candidate compound of the ABCA8 polypeptide.

The ABCA8 polypeptide in certain of the provided screening methods comprises the amino acid sequence of SEQ ID NO: 2, or at least residues 597-636 of SEQ ID NO: 2, or a sequence at least 98% identical to one of these sequences.

According to another aspect, the cells used in the screening method described above may be cells not naturally expressing, or alternatively expressing at a low level, the ABCA8 polypeptide, the cells being transfected with a recombinant vector according to the disclosure capable of directing the expression of a nucleic acid encoding the ABCA8 polypeptide.

Alternatively, cells may be used that have a natural deficiency in anion transport, or are pretreated with one or more anion channel inhibitors such as Verapamil™ or tetraethylammonium.

The anion used may be a radioactively labeled iodide, such as the salts $K^{125}I$ or $Na^{125}I$. Via this method, it is possible to establish a kinetic measurement of anion efflux during the experiment by measurement of efflux of the labeled anion periodically over time. The value of efflux of the labeled anion may be determined by measuring the quantity of labeled anion present at a given time in the cell culture supernatant. Further, the value of efflux of the labeled anion may be determined as the proportion of radioactivity found in the cell culture supernatant relative to the total radioactivity corresponding to the sum of the radioactivity found in the cell lysate and the radioactivity found in the cell culture supernatant.

Also provided are methods of screening a compound or small molecule active on the metabolism of lipophilic substances, an agonist or antagonist of the ABCA8 polypeptide, the methods including the following:

a) culturing cells of a human myocyte line in an appropriate culture medium, in the presence of purified human albumin;

b) incubating the cells of a) simultaneously in the presence of a compound stimulating the production of interleukin and of an agonist or antagonist candidate compound;

c) incubating the cells obtained in b) in the presence of an appropriate concentration of ATP;

d) measuring interleukin released into the cell culture supernatant; and

e) comparing the value of the release of the interleukin obtained in d) with the value of the interleukin released into the culture supernatant of cells which have not been previously incubated in the presence of the agonist or antagonist candidate compound.

According to a first aspect of the screening method described above, the cells used belong to the human or mouse myocytes. The compound used to stimulate the production of interleukin may a lipopolysaccharide. The production of all interleukins and TNF alpha by the cells and the level of expression of the messenger RNA encoding interleukin may be qualitatively and/or quantitatively determined.

XVII. Kits

Kits are provided which contain the necessary reagents for determining *ABCA8* gene copy number, for determining altered expression of *ABCA8* mRNA or ABCA8 protein, or for detecting polymorphisms in *ABCA8* alleles. Instructions provided in the diagnostic kits can include calibration curves, diagrams, illustrations, or charts or the like to compare with the determined (*e.g.*, experimentally measured) values or other results.

Kits are also provided that contain cells that serve as either positive or negative controls. These control cells can be compared to experimental samples containing similar cells, for instance cells of unknown gene activity, mutational state, protein expression level, and so forth.

A. ***Kits for Detection of ABCA8 Genomic Sequences***

The nucleotide sequences disclosed herein, and fragments thereof, can be supplied in the form of a kit for use in detection of *ABCA8* genomic sequences, for instance in order to diagnose a deficiency in transporter activity. In one embodiment of such a kit, an appropriate amount of one or more of the *ABCA8*-specific oligonucleotide primers is provided in one or more containers. In other embodiments, the oligonucleotide primers may be provided suspended in an aqueous solution or as a freeze-dried or lyophilized powder, for instance. The container(s) in which the oligonucleotide(s) are supplied can be any conventional container that is capable of holding the supplied form, for instance, microfuge tubes, ampoules, or bottles. In other embodiments, pairs of primers may be provided in pre-measured single use amounts in individual, typically disposable, tubes or equivalent containers. In one specific, non-limiting embodiment, the sample to be tested for the presence of *ABCA8* genomic amplification can be added to the individual tubes and *in vitro* amplification carried out directly.

The amount of each oligonucleotide primer supplied in the kit can be any appropriate amount, depending for instance on the market to which the product is directed. In one embodiment, the kit is adapted for research or clinical use and the amount of each oligonucleotide primer provided is an amount sufficient to prime several *in vitro* amplification reactions. Those of ordinary skill in the art know the amount of oligonucleotide primer that is appropriate for use in a single amplification reaction. General guidelines may for instance be found in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990), Sambrook *et al.* (In *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor, New York, 1989), and Ausubel *et al.* (In *Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998).

In one embodiment, a kit may include more than two primers, in order to facilitate the PCR *in vitro* amplification of *ABCA8* sequences, for instance the *ABCA8* gene, specific exon(s) or other portions of the gene, or the 5' or 3' flanking region thereof. By way of example, a kit may include primers for amplification of a nucleotide binding domain (e.g. residues 502-656 or 1308-1621 of SEQ ID NO: 2).

In some embodiments, kits may also include the reagents necessary to carry out PCR *in vitro* amplification reactions, including, for instance, DNA sample preparation reagents, appropriate buffers (*e.g.*, polymerase buffer), salts (*e.g.*, magnesium chloride), and deoxyribonucleotides (dNTPs). Instructions may also be included.

In other embodiments, kits may include either labeled or unlabeled oligonucleotide probes for use in detection of the *in vitro* amplified ABCA8 sequences. In one specific, non-limiting embodiment, the appropriate sequences for such a probe will be any sequence that falls between the annealing sites of the two provided oligonucleotide primers, such that the sequence the probe is complementary to is amplified during the *in vitro* amplification reaction.

In yet another embodiment, the kit provides one or more control sequences for use in the amplification reactions. The design of appropriate positive control sequences is well known to one of ordinary skill in the appropriate art.

B. ***Kits for Detection of ABCA8 mRNA Expression***

Kits similar to those disclosed above for the detection of *ABCA8* genomic sequences can be used to detect *ABCA8* mRNA expression levels. One embodiment of such a kit includes an appropriate amount of one or more of the oligonucleotide primers for use in reverse transcription amplification reactions, similarly to those provided above, with art-obvious modifications for use with RNA.

In some embodiments, kits for detection of *ABCA8* mRNA expression levels may also include the reagents necessary to carry out RT-PCR *in vitro* amplification reactions, including, for instance, RNA sample preparation reagents (including *e.g.*, an RNAse inhibitor), appropriate buffers (*e.g.*, polymerase buffer), salts (*e.g.*, magnesium chloride), and deoxyribonucleotides (dNTPs). Instructions also may be included.

In other embodiments, kits may include either labeled or unlabeled oligonucleotide probes for use in detection of the *in vitro* amplified target sequences. In one specific, non-limiting embodiment, the appropriate sequences for such a probe will be any sequence that falls between the annealing sites of the two provided oligonucleotide primers, such that the sequence the probe is complementary to is amplified during the PCR reaction.

In another embodiment, the kit provides one or more control sequences for use in the RT-PCR reactions. The design of appropriate positive control sequences is well known to one of ordinary skill in the appropriate art.

In yet other embodiments, kits may be provided with the necessary reagents to carry out quantitative or semi-quantitative Northern analysis of *ABCA8* mRNA. Such kits include, for instance, at least one *ABCA8*-specific oligonucleotide for use as a probe. This oligonucleotide may be labeled in any conventional way, including with a selected radioactive isotope, enzyme substrate, co-factor, ligand, chemiluminescent or fluorescent agent, hapten, or enzyme.

C. ***Kits For Detection of ABCA8 Protein or Peptide Expression***

In some embodiments, kits for the detection of ABCA8 protein expression include for instance at least one target protein specific binding agent (*e.g.*, a polyclonal or monoclonal antibody or antibody fragment) and may include at least one control. In another embodiment, the ABCA8 protein specific binding agent and control may be contained in separate containers. In other embodiments, the kits may also include means for detecting ABCA8:agent complexes, for instance the agent may be detectably labeled. If the detectable agent is not labeled, it may be detected by second antibodies, or protein A for example, which may also be provided in some kits in one or more separate containers. Such techniques are well known.

In another embodiment, the kits include instructions for carrying out the assay. Instructions will allow the tester to determine whether ABCA8 expression levels are altered, for instance in comparison to a control sample. In other embodiments, reaction vessels and auxiliary reagents such

as cells, chromogens, buffers, media, enzymes, etc. also may be included in the kits.

In one specific, non-limiting embodiment, an effective and convenient immunoassay kit such as an enzyme-linked immunosorbant assay can be constructed to test anti-ABCA8 antibody in human serum, as reported for detection of non-specific anti-ovarian antibodies (Wheatcroft *et al, Clin. Exp. Immunol.* 96:122-128, 1994; Wheatcroft *et al, Hum. Reprod.* 12:2617-2622, 1997). In one embodiment, expression vectors can be constructed using the human ABCA8 cDNA to produce the recombinant human ABCA8 protein in either bacteria or baculovirus (as described herein). In another embodiment, affinity purification is used to generate unlimited amounts of pure recombinant ABCA8 protein.

In one embodiment, an assay kit could provide the recombinant protein as an antigen and enzyme-conjugated goat anti-human IgG as a second antibody as well as the enzymatic substrates. Such kits can be used to test if the sera from a subject contain antibodies against ABCA8.

D. ***Kits for Detection of Homozygous versus Heterozygous Allelism***

Also provided are kits that allow differentiation between individuals who are homozygous versus heterozygous for a polymorphism of *ABCA8*. In one embodiment such kits provide the materials necessary to perform oligonucleotide ligation assays (OLA), for instance as described at Nickerson *et al.* (*Proc. Natl. Acad. Sci. USA* 87:8923-8927, 1990). In specific embodiments, these kits contain one or more microtiter plate assays, designed to detect allelism in the *ABCA8* sequence of a subject, as described herein.

In one embodiment, additional components in some of these kits may include instructions for carrying out the assay. Instructions will allow the tester to determine whether an *ABCA8* allele is homozygous or heterozygous. In other embodiments, reaction vessels and auxiliary reagents such as chromogens, buffers, enzymes, etc. may also be included in the kits.

In another embodiment, the kit may provide one or more control sequences for use in the OLA reactions. The design of appropriate positive control sequences is well known to one of ordinary skill in the appropriate art.

E. ***Kits for Identifying Modulators of ABCA8 Activity***

Also provided are kits that allow for the identification of modulators of ABCA8 activity. In one embodiment, such kits provide the materials necessary to assess the activity of ABCA8 *in vitro*. In one embodiment, this kit contains aliquots of isolated ABCA8 and cultured cells. In another embodiment, the kit contains cell lines that express either wildtype or mutant ABCA8. In yet another embodiment, additional components in some of these kits may include instructions for carrying out the assay. In other embodiments, reaction vessels and auxiliary reagents such as chromogens, buffers, media, enzymes, etc. may also be included in the kits.

The disclosure is illustrated by the following non-limiting Examples.

EXAMPLES

EXAMPLE 1: Identification of human ABCA8 in genomic databases

This example provides a description of how the *ABCA8* genomic sequence was first identified.

EST / genomic and protein database searches

Expressed sequence tags (EST) of ABCA1-like genes as described by Allikmets *et al.* (*Hum Mol Genet,* 5, 1649-1655, 1996) were used to search Genbank and UniGene nucleotide sequence databases using BLAST2 (Altschul *et al., Nucleic Acids Res.*, 25:3389-3402, 1997). The main protein sequences databases screened were Swissprot, TrEMBL, Genpept and PIR. The searches were performed between December 1999 and April 2000.

Sequence analysis

The genomic DNA analysis was performed by combination of several gene-finding programs such as GENSCAN (Burge and Karlin, *J Mol Biol.;* 268(1):78-94, 1997), FGENEH/FEXH (Solovyev and Salamov, *ISMB*; 5:294-302, 1997), and XPOUND (Thomas and Skolnick, *J Math Appl Med Biol.*;11(1):1-16, 1994). The combination of different tools yield increased sensitivity and specificity. The second step in the genomic DNA analysis was homology searching in the EST and protein databases. Combination of software performing database searching and software for exon/intron prediction gave the most sensitive and specific results. Sequence assembly and analysis were performed using the Genetics Computer Group (GCG) sequence analysis software package.

Multiple alignments were generated by GAP software from GCG package and the Dialign2 program (Morgenstern *et al., Proc Natl Acad Sci USA.*; 93(22): 12098-103, 1996), the FASTA3 package (Pearson and Lipman, *Proc Natl Acad Sci USA.*; 85(8): 2444-8, 1988) and SIM4 (Florea *et al., Genome Res.* 1998 Sep;8(9):967-74, 1998). The specific ABCA motifs used to search the data bases were the TMN, TMC, NBF1 and NBF2 described in the literature (Broccardo *et al., Biochim. Biophys. Acta* 1461(2):395-404, 1999). This corresponds in ABCA1 to residues 630-846 for the N terminal and from 1647-1877 for the C terminal set of membrane spanners. The NBD corresponds to the extended nucleotide binding domain, *i.e.* in ABCA1 it spans from amino acids 885-1152 for the N-terminal one and 1918-2132 for the C-terminal one.

The alignments of the TMN and TMC domains of the 17q24 genes were used to perform phylogenetic analysis along with ABCA1, ABCA2, ABCA3, ABCA4, and *C. elegans* ABCA-related gene. The sequences were aligned with PILEUP (Wisconsin Package) and converted to a PHYLIP format. Phylogenetic analysis was performed with the PHYLIP Protpars (maximum parsimony) and Protdist (neighbor joining) methods (PHYLIP package). The Seqboot program was used for

bootstrap analysis (100 iterations) and consensus trees were generated with Consense.

Gene prediction analysis

Contigs assembled from BAC AC007763, AC005495, and AC005922 from 17q24 were analyzed using the Affy Genes, Ensembl Genes, and Fgenesh++ Genes dropdown combo box on the 'dense' setting at the University of California at Santa Cruz web-site. A parallel analysis was carried forward using the Celera Discovery System with their proprietary gene prediction software.

Results

Using BAC sequences AC005495 and AC005922, an electronic intron/exon prediction was performed. The ABCA8 gene sequence was partially contained in the contig of these BACs as the 5' end (Figure 1).

Phylogenetic analysis of the ATP-binding domains of the ABCA genes located in this chromosomal region demonstrates that the N- and C-terminal domains form separate branches (Figure 3). The C-terminal ATP-binding domains of the 17q24 genes are more closely related to the C-terminal domains of the other ABCA1-like genes than to the N-terminal domains of the same proteins. Thus the entire ABCA1 subfamily appears to have arisen from a single ancestral full transporter gene. Genes in the 17q24 cluster form a distinct group within the ABCA subfamily.

EXAMPLE 2: Human *ABCA8* cDNA.

This Example describes the isolation and identification of cDNA molecules encoding the full-length human ABCA8 protein.

Reverse transcription

In a total volume of 11.5 µl, 500 ng of mRNA poly(A)+ (prepared using a CLONTECH mRNA purification kit) mixed with 500 ng of oligo-dT were denatured at 70°C for 10 minutes and then chilled on ice. After addition of 10 units of RNasin, 10 mM DTT, 0.5 mM dNTP, Superscript first strand buffer and 200 units of Superscript II (Life Technologies), the reaction was incubated for 45 minutes at 42°C.

PCR

Each polymerase chain reaction contained 400 µM each dNTP, 2 units of *Thermus aquaticus* (Taq) DNA polymerase (Ampli Taq Gold; Perkin Elmer), 0.5 µM each primer, 2.5 mM $MgCl_2$, PCR buffer and 50 ng of DNA, or about 25 ng of cDNA, or $1/50^{th}$ of a primary PCR mixture. Reactions were carried out for 30 cycles in a Perkin Elmer 9700 thermal cycler in 96-well microtiter plates. After an initial denaturation at 94°C for 10 minutes, each cycle consisted of: a denaturation step of 30 seconds (94°C), a hybridization step of 30 seconds (64°C for 2 cycles, 61°C for 2 cycles, 58°C for 2

cycles and 55°C for 28 cycles), and an elongation step of 1 minute/kb (72°C). PCR amplification ended with a final 72°C extension of 7 minutes. In case of RT-PCR, control reactions without reverse transcriptase and reactions containing water instead of cDNA were performed for every sample.

Primers

Oligonucleotides were selected using Prime from GCG package or Oligo 4 (National Biosciences, Inc.) softwares. Primers were ordered from Life Technologies, Ltd. and used without further purification.

DNA Sequencing

PCR products were analyzed and quantified by agarose gel electrophoresis, purified with a P100 column. Purified PCR products were sequenced using ABI Prism BigDye terminator cycle sequencing kit (Perkin Elmer Applied Biosystems). The sequence reaction mixture was purified using Microcon-100 microconcentrators (Amicon, Inc., Beverly). Sequencing reactions were resolved on an ABI 377 DNA sequencer (Perkin Elmer Applied Biosystems) according to manufacturer's protocol (Applied Biosystems, Perkin Elmer).

Results

The complete sequence of the human ABCA8 cDNA is shown in SEQ ID NO: 1.

EXAMPLE 3: Method of Making *ABCA8* Encoding Sequences

The foregoing discussion describes the original means by which the *ABCA8* cDNA was obtained and also provides the nucleotide sequence of this clone. With the provision of this sequence information, the polymerase chain reaction (PCR) or other similar amplification techniques may now be utilized in a more direct and simple method for producing *ABCA8* encoding sequences.

Total RNA is extracted from human cells by any one of a variety of methods well known to those of ordinary skill in the art. Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor, New York, 1989, and Ausubel *et al. Current Protocols in Molecular Biology*, John Wiley & Sons, New York, 1998 provide descriptions of methods for RNA isolation. Any human cell line derived from a nonABCA8 deleted individual would be suitable, such as the widely used HeLa cell line, or the WI-38 human skin fibroblast cell line available from the American Type Culture Collection, Rockville, MD. The extracted RNA is then used as a template for performing the reverse transcription-polymerase chain reaction (RT-PCR) amplification of cDNA. Methods and conditions for RT-PCR are described in Kawasaki *et al.*, (In *PCR Protocols, A Guide to Methods and Applications*, Innis *et al.* (eds.), 21-27, Academic Press, Inc., San Diego, California, 1990). The selection of PCR primers will be made according to the portions of the cDNA that are to be amplified. Primers may be chosen to amplify small segments of a cDNA or the entire cDNA molecule. Variations in amplification conditions may be required to accommodate primers and amplicons of

differing lengths and composition; such considerations are well known in the art and are discussed in Innis *et al.* (*PCR Protocols, A Guide to Methods and Applications*, Academic Press, Inc., San Diego, CA, 1990). By way of example, *ABCA8*-encoding sequences may be amplified theoretically using the following combination of 5' → 3' primers:

Full-length SEQ ID NO:1:
primer 1: 5' CAGAAGTCAAATAG: 3' (SEQ ID NO: 3)
primer 2 5' CAAAACTATATATATG: 3' (SEQ ID NO: 4).

An additional set of primers that may be used to amplify an ABCA8 encoding sequence are:

primer 3: 5' CATGGTAGAGACCAGGAGAAAA 3' (SEQ ID NO: 9)
primer 4: 5' CTCAGTCATGTGAGCTGTTGC 3' (SEQ ID NO: 10)

The coding sequence of *ABCA8* may be amplified using the following combination of primers:

primer 5 5' ATGAGGAAGAGAAAG 3' (SEQ ID NO: 5)
primer 6 5' TTAAGGCTCTTCCTGGGGG 3' (SEQ ID NO: 6).

The newly identified portion of the *ABCA8* cDNA (nucleotides 1927-2046 of SEQ ID NO: 1) may be amplified using the following combination of primers:

primer 7 5' ATACAAAGGGTTC 3' (SEQ ID NO: 7)
primer 8 5' CTGAGGATCTCC 3' (SEQ ID NO: 8).

These primers are illustrative only; one skilled in the art will appreciate that many different primers may be derived from the provided cDNA sequence in order to amplify particular regions thereof.

EXAMPLE 4: Tissue Distribution of *ABCA8* Transcripts.

The profile of expression of *ABCA8* polynucleotides can be determined using art-known techniques, such as PCR-coupled reverse transcription, which has been described for instance by Sambrook *et al.* (1989, *Molecular cloning: A laboratory manual.* 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). This example provides representative methods and results from various tissues.

RT-PCR

For example, in the case of an analysis by reverse transcription, a pair of primers as described above (*e.g.*, SEQ ID NO: 3 and 4) may be synthesized corresponding to a human *ABCA8* cDNA (SEQ ID NO: 1). This primer pair may be used to detect the corresponding ABCA8 cDNA.

The polymerase chain reaction (PCR) was carried out on commercially provided first-strand cDNA templates (CLONTECH) corresponding to retro-transcribed polyA$^+$ mRNAs. Reverse transcription to cDNA was carried out with the enzyme SUPERSCRIPT II (GibcoBRL, Life Technologies) according to the conditions described by the manufacturer.

The polymerase chain reaction was carried out according to standard conditions, in 20 µl of reaction mixture with 25 ng of cDNA preparation. The reaction mixture was composed of 400 µM of each of the dNTPs, 2 units of Thermus aquaticus (Taq) DNA polymerase (Ampli Taq Gold; Perkin Elmer), 0.5 µM of each primer, 2.5 mM $MgCl_2$, and PCR buffer. Thirty four PCR cycles [denaturing 30 seconds at 94°C, annealing of 30 seconds divided up as follows during the 34 cycles: 64°C (2 cycles), 61°C (2 cycles), 58°C (2 cycles), and 55°C (28 cycles), and an extension of one minute per kilobase at 72°C] were carried out after a first step of denaturing at 94°C for 10 minutes using a Perkin Elmer 9700 thermocycler. The PCR reactions were visualized on agarose gel by electrophoresis, and their relative level quantified.

Results

Using RT-PCR, the amount of ABCA8 transcript was measured in various tissues. The results are shown in Table 7.

Table 7: Relative Level of ABCA8 Transcript in Human Tissues

TISSUE	*ABCA8*
brain	2*
heart	**13**
kidney	1.8
liver	**21**
lung	4.2
pancreas	<1.0
placenta	1.7
skeletal muscle	<1.0
colon	6
ovary	**98**
peripheral blood lymphocytes	<1.0
prostate	6.1
small intestine	9
spleen	5.5
testis	**45**
thymus	1.3

* - units of measurement are ng of DNA as compared to a standard curve

cDNA fragments obtained by PCR may be used as probes for a Northern blot analysis (see below), or can be sequenced in order to determine the exact nucleotide sequence of the amplified nucleic acids.

EXAMPLE 5: Northern Blot Analysis of ABCA8 Transcript.

The profile of expression of *ABCA8* polynucleotides can be determined using art-known techniques, such as Northern blot analysis, which has been described for instance by Sambrook *et al.* (1989, *Molecular cloning: a laboratory manual.* 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York). This example provides representative methods.

Preparation of probe

PCR products are gel-purified using Qiaquick® column (Qiagen). 10-20 ng of purified PCR product are radiolabeled with [γ^{32}P]dCTP (Amersham; 6000 Ci/mmol, 10 mCi/ml) by the random priming method (Rediprime kit; Amersham) according to the manufacturer's protocol. Unincorporated radioactive nucleotides are separated from the labeled probe by filtration on a G50 microcolumn (Pharmacia). Probe is competed with 50 µg of denatured human Cot1 DNA for 2 hours at 65°C.

Hybridization

Prehybridization of the Northern blot (6 hours at 42°C) with hybridization solution (5x SSPE, 5x Denhardt's, 2.5% Dextran, 0.5% SDS, 50% formamide, 100 µg/ml denatured salmon sperm DNA, 40 µg denatured human DNA) is followed by hybridization with radiolabeled probe (2.10^6 cpm/ml hybridization solution, as described above) and 40 µg of denatured human DNA. Filters are washed in 2x SSC for 30 minutes at room temperature, twice in 2x SSC-0.1% SDS for 10 minutes at 65°C and twice in 1x SSC-0.1% SDS for 10 minutes at 65°C. Northern blots are analyzed after overnight (or other appropriate duration) exposure on, for instance, a Storm system (Molecular Dynamics, Sunnyvale, CA). Human transferrin probe can be used to control the amount of RNA loaded in each lane of the membrane.

EXAMPLE 6: Production of Normal and Mutated ABCA8 Polypeptides.

The normal ABCA8 polypeptide (SEQ ID NO: 2) encoded by complete corresponding cDNA (SEQ ID NO: 1) whose isolation is described in Example 2, or a mutated ABCA8 polypeptides whose complete cDNA may also be obtained according to the techniques described in Example 2, may be produced in a bacterial or insect cell expression system using the baculovirus vectors or in mammalian cells with or without the vaccinia virus vectors. All the methods are now widely described and are known to persons skilled in the art. A detailed description thereof will be found for example in Ausubel *et al.* (1989, *Current Protocols in Molecular Biology*, Green Publishing Associates and Wiley Interscience, N.Y).

EXAMPLE 7: Determination of Polymorphisms/Mutations in *ABCA8*.

The detection of polymorphisms or of mutations in the sequences of the transcripts or in the genomic sequence of *ABCA8* may be carried out according to various protocols. One method that may be used is direct sequencing.

For subjects from whom it is possible to obtain an mRNA preparation, the method consists of preparing the cDNAs and sequencing them directly. For subjects for whom only DNA is available, and in the case of a transcript where the structure of the corresponding gene is unknown or partially known, it may be beneficial to determine the intron-exon structure of the transcript as well as the genomic sequence of the corresponding gene. This involves, in a first instance, isolating the genomic DNA BAC or cosmid clone(s) corresponding to the transcript studied, sequencing the insert of the corresponding clone(s), and determining the intron-exon structure by comparing the cDNA sequence to that of the genomic DNA obtained.

The technique of detection of mutations by direct sequencing consists of comparing the genomic sequences of ABCA8 obtained from homozygotes for the disease or from at least 8 individuals (4 individuals affected by the pathology studied and 4 individuals not affected) or from at least 32 unrelated individuals from the studied population. Sequence divergences constitute polymorphisms. All modifications of the amino acid sequence of the wild-type protein may be mutations capable of affecting the function of a protein, which may yield information regarding the mutation(s) of the disease (denoted genotype-phenotype correlation) in the pedigree, or of a pharmacological response to a therapeutic molecule in the pharmacogenomic studies.

EXAMPLE 8: Identification of a Causal Gene for a Disease Linked to ABCA8

Among the mutations identified according to the method described in Example 7, all those associated with the disease phenotype are capable of being causal. These results are validated by sequencing the gene in the affected individuals and their relations (whose DNA is available).

Moreover, Northern blot or RT-PCR analysis, according to the methods described herein, using RNA specific to affected or non-affected individuals makes it possible to detect notable variations in the level of expression of the gene studied, in particular the absence of transcription of the gene.

EXAMPLE 9: Validation of the Expression of the Human ABCA8 cDNA:

Polyclonal antibodies specific for a human ABCA8 polypeptide may be prepared as described herein in rabbits and chicks by injecting a synthetic polypeptide fragment derived from an ABCA8 protein, including all or part of an amino acid sequence as shown in SEQ ID NO: 2, for instance a fragment chosen from within residues 597-636 of SEQ ID NO: 2. These polyclonal antibodies are used to detect and/or quantify the expression of the ABCA8 gene in cells and animal models by immunoblotting and/or immunodetection.

By way of example, a biological activity of ABCA8 may be monitored by quantifying the cholesterol fluxes induced by apoA-I using cells transfected with the vector pCMV-ABCI which have been loaded with cholesterol (Remaley *et al.*, 1997, *ATVB*, 17:1813).

EXAMPLE 10: Expression *in vivo* of the *ABCA8* Gene in Various Animal Models:

An appropriate volume (*e.g.*, 100 to 300 µl) of a medium containing the purified recombinant adenovirus (pABCA-AdV or pLucif-AdV) containing from 10^8 to 10^9 lysis plaque-forming units (pfu) are infused into the Saphenous vein of mice (C57BL/6, both control mice and models of transgenic or knock-out mice) on day 0 of the experiment.

The evaluation of the physiological role of the ABCA8 protein in the transport of cholesterol or inflammatory lipid substances is carried out by determining the total quantity of cholesterol or appropriate inflammatory lipid substances before (day zero) and after (days 2, 4, 7, 10, 14) the administration of the adenovirus.

Kinetic studies with the aid of radioactively labeled products are carried out on day 5 after the administration of the vectors rLucif-AdV and rABCA-AdV in order to evaluate the effect of the expression or function of ABCA8 on the transport of cholesterol and inflammatory lipid substances.

Furthermore, transgenic mice and rabbits over-expressing gene may be produced, in accordance with the teaching of Vaisman (*J. Biol. Chem.* 270(20): 12269-12275, 1995) and Hoeg (*Proc. Nat. Acad. Sci. USA.* 93(21):11448-11453, 1996) using constructs containing the human ABCA8 cDNA under the control of endogenous promoters such as ABCA8 or CMV or apoE.

The evaluation of the long-term effect of the expression of ABCA8 on the kinetics of the lipids involved in the mediation of the inflammation may be carried out as described above.

An appropriate volume (*e.g.*, 100 to 300 µl) of a medium containing the purified recombinant adenovirus (pABCA-AdV or pLucif-AdV) containing from 10^8 to 10^9 lysis plaque-forming units (pfu) are infused into the Saphenous vein of mice (C57BL/6, both control mice and models of transgenic or knock-out mice) on day 0 of the experiment.

The evaluation of the physiological role of the ABCA8 protein in the transport of cholesterol or inflammatory lipid substances is carried out by determining the total quantity of cholesterol or appropriate inflammatory lipid substances before (day zero) and after (days 2, 4, 7, 10, 14) the administration of the adenovirus.

Kinetic studies with the aid of radioactively labeled products are carried out on day 5 after the administration of the vectors rLucif-AdV and rABCA-AdV in order to evaluate the effect of the expression of ABCA8 on the transport of cholesterol and inflammatory lipid substances.

Furthermore, transgenic mice and rabbits over-expressing gene may be produced, in accordance with the teaching of Vaisman (*J. Biol. Chem.* 270(20):12269-12275, 1995) and Hoeg (*Proc. Nat. Acad. Sci. USA.* 93(21):11448-11453 1996) using constructs containing the human ABCA8 cDNA under the control of endogenous promoters such as ABCA8 or CMV or apoE.

The evaluation of the long-term effect of the expression of ABCA8 on the kinetics of the lipids involved in the mediation of the inflammation may be carried out as described above.

The present disclosure is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the disclosure in addition to those described herein will become apparent to those skilled in the art from the foregoing description and the accompanying Figures. Such modifications are intended to fall within the scope of the appended claims.

EXAMPLE 11: Sequence Analysis of Human *ABCA8* cDNA.

This example provides a description of how the *ABCA8* sequence was analyzed.

Sequence assembly and analysis was performed using the Autoassembler software (ABI Prism, Perkin Elmer Applied Biosystems, Palo Alto, CA), the Genetics Computer Group (GCG) (Accelrys Corp., San Diego, CA) sequence analysis software package, Autoassembler software (ABI Prism, Perkin Elmer Applied Biosystems, Palo Alto, CA) and the LFASTA program package (*see* Pearson and Lipman, *Proc. Natl. Acad. Sci. U.S.A.* 85(8): 2444-2448, 1988). The programs were operated using default parameters, and were used to assemble and generate the consensus sequence for ABCA8.

Sequence analysis of this PCR product revealed that the two putative NBDs are separated by a putative transmembrane domain. The order and sequence of these 3 putative protein motifs are indeed characteristic of an ABCA gene.

EXAMPLE 12: Phylogenetic analysis of the relationship between the ABCA8 to other ABCA transporters.

This example provides a description of how the evolutionary relationship between human *ABCA8* and other ABCA transporters was determined.

Phylogenetic analysis was performed using the ATP-binding domain, which is conserved across the species and in different subfamilies without large deletions or insertions (see Bodenmiller *et al., DNA Seq.* 13(2): 77-83, 2002). Analysis was conducted using both ATP-binding domains, as all twelve human ABCA subfamily members known to date contain two such domains.

Phylogenetic analysis indicated that ABCA8 belongs to the subgroup inside the A subfamily, which contains the best-characterized members of this subfamily, ABCA1 (see Tanaka *et al.*, *Biochem. Biophys. Res. Commun.* 25;283(5): 1019-1025, 2001) and ABCA4 (see Bungert *et al., J. Biol. Chem.* 276(26): 23539-23546, 2001). The second subgroup includes five genes that form a tandem cluster at chromosome 17q24. When compared to other ABCA proteins, ABCA8 has a short (about 90 amino acids) region of similarity at the N-terminus, and larger region of similarity starting from amino acid 3070. This suggests that the large size of ABCA8 is due to the large insertion between these positions at the first extracellular domain.

Embodiments of this disclosure provide several ABCA8 proteins and nucleic acid molecules, and methods of isolating, making, and using these molecules. Further embodiments provide methods

for ameliorating, treating, detecting, prognosing and diagnosing diseases related to expression of ABCA8. It will be apparent that the precise details of the methods described may be varied or modified without departing from the spirit of the described invention. We claim all such modifications and variations that fall within the scope and spirit of the claims below.

5

CLAIMS

1. A substantially purified human ABCA8 protein, comprising at least one amino acid sequence as set forth in

(a) SEQ ID NO: 2;

(b) residues 597-636 of SEQ ID NO: 2;

(c) sequences having 98% sequence identity to (a) or (b); or

(d) a conservative variant of (a).

2. The substantially purified human ABCA8 protein of claim 1, which protein has ABCA8 biological activity.

3. The substantially purified human ABCA8 protein of claim 1, wherein the protein is an active transporter protein associated with transport of a lipophilic compound.

4. The substantially purified human ABCA8 protein of claim 1, comprising the amino acid sequences as set forth in SEQ ID NO: 2 or residues 597-636 of SEQ ID NO: 2.

5. The substantially purified human ABCA8 protein of claim 1, comprising an amino acid sequence having 98% sequence identity to SEQ ID NO: 2.

6. The substantially purified human ABCA8 protein of claim 1, comprising an amino acid sequence having 90% sequence identity to residues 597-636 of SEQ ID NO: 2 but containing all of SEQ ID NO: 2.

7. A substantially purified functional human ABCA8 protein consisting essentially of the amino acid sequence as set forth in SEQ ID NO: 2.

8. A substantially purified nucleic acid molecule encoding the protein according to claim 1.

9. The substantially purified nucleic acid molecule of claim 8, comprising a nucleotide sequence as set forth in:

(a) SEQ ID NO: 1;

(b) nucleotides 1927-2046 of SEQ ID NO: 1; or

(c) sequences having at least 98% sequence identity with (a) or (b).

10. The substantially purified nucleic acid molecule of claim 8, comprising nucleotide sequences as set forth in SEQ ID NO: 1 and nucleotides 1927-2046 of SEQ ID NO: 1.

11. The substantially purified nucleic acid molecule of claim 8, comprising a nucleotide sequence having at least 98% sequence identity with SEQ ID NO: 1 or nucleotides 1927-2046 of SEQ ID NO: 1.

12. A recombinant nucleic acid molecule comprising a promoter sequence operably linked to the nucleic acid molecule according to claim 8.

13. A recombinant nucleic acid molecule according to claim 12, wherein the recombinant polynucleotide is in antisense orientation relative to the promoter sequence.

14. A recombinant vector comprising the recombinant nucleic acid molecule of claim 12.

15. A cell transformed with a recombinant nucleic acid molecule according to claim 12.

16. A transgenic, non-human animal, comprising the cell of claim 14.

17. An isolated nucleic acid molecule according to claim 8, wherein the molecule hybridizes with a nucleic acid probe comprising the sequence shown in SEQ ID NO: 1 or nucleotides 1927-2046 of SEQ ID NO: 1 under wash conditions of 2X SSC and 0.1% SDS at 65°C.

18. An isolated nucleic acid molecule according to claim 8, wherein the molecule hybridizes with a nucleic acid probe comprising the sequence shown in SEQ ID NO: 1 or nucleotides 1927-2046 of SEQ ID NO: 1, under wash conditions of 0.1X SSC and 0.1% SDS at 65°C.

19. A specific binding agent that specifically binds an epitope of the protein encoded by the protein of claim 1.

20. The binding agent of claim 19, wherein the agent specifically binds an epitope of the protein encoded by the amino acid sequence as set forth in SEQ ID NO: 2 or residues 597-636 of SEQ ID NO: 2.

21. The specific binding agent of claim 19, wherein the agent is an antibody.

22. A method of detecting a biological condition of a subject associated with an altered expression of an ABCA8 nucleic acid, comprising detecting

(1) a mutation in the *ABCA8* nucleic acid that is associated with disease;

(2) altered expression of an *ABCA8* nucleic acid; or

(3) expression of a mutant *ABCA8* nucleic acid in a biological sample, wherein the mutation, altered expression, or expression of the mutant *ABCA8* nucleic acid indicates that the subject has the biological condition.

23. The method of claim 22, wherein the method comprises evaluating defective extra- or intra-cellular transport.

24. The method of claim 22, wherein the altered expression of the *ABCA8* nucleic acid comprises an alteration in a cellular level of *ABCA8* nucleic acid or ABCA8 protein, in comparison to the level detected in a control sample.

25. The method of claim 22, wherein detected altered expression of the *ABCA8* nucleic acid comprises Southern blot analysis, quantitative polymerase chain reaction or semi-quantitative polymerase chain reaction.

26. The method of claim 22, wherein detecting altered expression of the *ABCA8* nucleic acid comprises sequencing, chemical cleavage, denaturing gradient gel electrophoresis, or hybridization with allele specific oligonucleotides.

27. The method of claim 22, wherein the biological condition comprises elevated cholesterol levels, neurological disease, retinal degeneration, drug resistance, or an increased susceptibility to heart disease.

28. The method of claim 22, wherein the method is used for detecting a cell that is resistant to chemotherapy.

29. The method of claim 22, wherein the method is used for detecting a predisposition to elevated cholesterol or heart disease or for presymptomatic screening of an individual for hypercholesterolemia or defects in extra- or intracellular transport of ions, amino acids, peptides, sugars, vitamins or steroid hormones.

30. The method of claim 22, wherein the altered expression of the *ABCA8* nucleic acid comprises an increased or decreased expression of ABCA8 in a subject as compared to a control.

31. The method of claim 22, comprising reacting at least one ABCA8 molecule contained in a biological sample from the subject with a reagent comprising an ABCA8 specific binding agent to form an ABCA8:agent complex.

32. The method of claim 31, wherein the ABCA8 molecule is encoded by a nucleic acid having a nucleotide sequence as shown in SEQ ID NO: 1.

33. The method of claim 31, wherein the ABCA8:agent complex is detected by nucleotide hybridization.

34. The method of claim 31, wherein the agent is a labeled nucleotide probe.

35. The method of claim 34, wherein the nucleotide probe has a sequence selected from the group consisting of:

(a) SEQ ID NO: 1;

(b) nucleotides 1927-2046 of SEQ ID NO: 1

(c) nucleic acid sequences having at least 98% sequence identity with (a) or (b); and

(e) fragments of (a) or (b) at least 15 nucleotides in length.

36. The nucleotide probe used in the method of claim 35.

37. The method of claim 31, wherein the ABCA8:agent complex is detected by Western blot assay.

38. The method of claim 37, wherein the Western blot assay uses an antibody generated against a nucleotide binding or transmembrane domain of the protein encoded by the amino acid set forth in SEQ ID NO: 2, or a conservative substitution thereof.

39. The method of claim 31, wherein the complexes are detected by ELISA.

40. The method of claim 38 or 39, wherein the antibody is a monoclonal antibody.

41. The antibody of claim 40, which recognizes an antigenic peptide comprising a sequence within residues 597-636 of SEQ ID NO: 2.

42. The method of claim 37, wherein the ABCA8 protein comprises a sequence selected from the group consisting of:

(a) SEQ ID NO: 2;

(b) residues 597-636 of SEQ ID NO: 2;

(c) amino acid sequences having at least 98% sequence identity with (a) or (b); and

(d) conservative variants of (a) or (b).

43. The method of claim 22, further comprising in vitro amplifying an ABCA8 nucleic acid using oligonucleotide primers prior to detecting altered expression.

44. The method of claim 43, wherein the *ABCA8* nucleic acid is in vitro amplified using at least one oligonucleotide primer having a sequence identical to at least 15 contiguous nucleotides from SEQ ID NO: 1 or nucleotides 1927-2046 of SEQ ID NO: 1.

45. An oligonucleotide primer used in the method of claim 43

46. A recombinant DNA vector comprising the oligonucleotide primer of claim 45

47. A recombinant nucleic acid molecule comprising a promoter sequence operably linked to the oligonucleotide primer sequence of claim 45

48. The recombinant nucleic acid molecule of claim 45wherein the oligonucleotide primer sequence is in antisense orientation relative to the promoter sequence.

49. A cell transformed with the recombinant nucleic acid molecule of claim 47

50. A transgenic non-human animal, comprising the recombinant nucleic acid molecule of claim 48.

51. A method of screening for an agent that modulates ABCA8 transporter activity, the method comprising:

transfecting a cell with an expression vector, wherein the expression vector comprises a nucleic acid molecule having a sequence as set forth in SEQ ID NO: 1, or a conservative variation thereof, operably linked to a promoter sequence;

contacting the cell with a test agent; and

detecting a change in the level of expression of the ABA13 protein, wherein a change in the level indicates that the test agent modulates the expression of the ABCA8 transporter protein.

52. A method of detecting an ABCA8 transporter protein in a biological sample, comprising:

amplifying nucleotide residues as set forth in SEQ ID NO: 1, or a conservative substitution thereof, with two or more oligonucleotide primers that specifically bind the nucleotide residues as set forth in SEQ ID NO: 1; and

detecting a level of an amplified product, thereby detecting the ABCA8 transporter protein.

53. The method of claim 43 wherein the sample comprises blood, a blood product, urine, saliva, a tissue biopsy, a surgical specimen, an amniocentesis sample, or autopsy material.

54. A kit for detecting an excess or deficiency of ABCA8 protein in a subject using the method of claim 38, the kit comprising an ABCA8 protein specific binding agent.

55. The kit of claim 54 wherein the agent specifically binds to an epitope within:

(a) the amino acid sequence shown in SEQ ID NO: 2;

(b) the amino acid sequence shown in residues 597-636 of SEQ ID NO: 2; or

(c) the amino acid sequence having at least 98% sequence identity to the sequence specific in (a) or (b).

56. The kit of claim 55 further comprising a means for detecting binding of the ABCA8 protein binding agent to an ABCA8 polypeptide.

57. The kit of claim 55 wherein the subject is a mammal.

58. The kit of claim 57 wherein the mammal is a human or mouse.

59. The kit of claim 55, wherein the abnormal ABCA8 expression results in altered extra- or intracellular transport.

60. The kit of claim 55, wherein the ABCA8 epitope binding agent is an antibody.

61. A kit for detection of a genetic mutation in a sample of nucleic acid for a subject, comprising:

(a) a first container containing an oligonucleotide capable of specifically hybridizing with an ABCA8 nucleic acid; and

(b) a second container containing a labeled nucleic acid probe that is fully complementary to the oligonucleotide.

62. The kit of claim 61, wherein the labeled nucleic acid probe has a length of between 5 and 500 nucleotides.

63. A kit for determining whether or not a subject has a biological condition associated with an abnormal ABCA8 expression by detecting altered expression of ABCA8 protein in a biological sample from a subject, comprising:

a container comprising an antibody specific for an epitope of the ABCA8 protein; and instructions for using the kit, the instructions indicating steps for:

performing a method to detect the presence of ABCA8 protein in the sample; and

analyzing data generated by the method,

wherein the instructions indicate that underabundance of ABCA8 protein in the sample indicates that the individual has or is predisposed to the biological condition.

64. The kit of claim 63, wherein the agent is capable of specifically binding to an epitope within

(a) the amino acid sequence shown in SEQ ID NO: 2;

(b) the amino acid sequence shown in amino acid residues 597-636 of SEQ ID NO: 2; or

(c) a conservative variant of (a) or (b).

65. The kit of claim 63 further comprising a container that comprises a detectable antibody that binds to the ABCA8 protein-specific antibody.

66. An in vitro assay kit for determining whether or not a subject has a biological condition associated with an abnormal ABCA8 expression, the kit comprising:

a container comprising an ABCA8 protein specific antibody;

a container comprising a negative control sample; and

instructions for using the kit, the instructions indicating steps for:

performing a test assay to detect a quantity of ABCA8 protein in a test sample of tissue and/or bodily fluid from the subject

performing a negative control assay to detect a quantity of ABCA8 protein in the negative control sample; and

comparing data generated by the test assay and negative control assay,

wherein the instructions indicate that a quantity of ABCA8 protein in the test sample less than the quantity of ABCA8 protein in the negative control sample indicates that the subject has the biological condition.

67. The kit of claim 66 further comprising a container that comprises a detectable antibody that binds to the ABCA8 protein specific antibody.

68. An in vitro amplification assay kit, comprising:

a first container comprising an in vitro amplification primer that specifically amplifies the nucleic acid that encodes ABCA8 protein;

a second container comprising a size marker, the size marker being the expected size of amplified DNA if the nucleic acid that encodes ABCA8 protein is present in the sample; and

instructions for using the kit, wherein the instructions indicate steps for performing a method to detect and/or quantify the nucleic acid that encodes ABCA8 protein in the sample; and

analyzing data generated by the method,

wherein the instructions indicate that an altered level of nucleic acid that encodes ABCA8 protein in the sample indicates that the individual has a biological condition associated with ABCA8 expression.

69. A method of modifying a level of expression of an ABCA8 protein in a subject, comprising expressing in the subject a recombinant genetic construct comprising a promoter operably linked to a nucleic acid molecule, wherein the nucleic acid molecule comprises at least 15 consecutive nucleotides of the sequence shown in SEQ ID NO: 1, or nucleotides 1927-2046 of SEQ ID NO: 1.

70. The method of claim 69, wherein the nucleic acid molecule is in antisense orientation compared to the promoter.

71. An isolated fragment of SEQ ID NO: 2 of at least 120 nucleotides in length, comprising at least one amino acid sequence as set forth in:

(a) residues 597-636 of SEQ ID NO: 2; or

(b) sequences having at least 98% sequence identity with (a); or

(c) conservative variants of (a).

72. The isolated fragment of claim 71, consisting essentially of a peptide having the amino acid sequence as set forth in SEQ ID NO: 2.

73. The method of claim 43, wherein the ABCA8 nucleic acid is in vitro amplified using at least one oligonucleotide primer derived from an ABCA8 encoding sequence.

74. The method of claim 73, wherein the oligonucleotide primer comprises a sequence selected from the group consisting of:

(a) SEQ ID NO: 3;

(b) SEQ ID NO: 4;
(c) SEQ ID NO: 5;
(d) SEQ ID NO: 6;
(e) SEQ ID NO: 7;
(f) SEQ ID NO: 8;
(g) SEQ ID NO: 9; and
(h) SEQ ID NO: 10.

75. A pharmaceutical composition, comprising a polypeptide having the sequence set forth in SEQ ID NO: 2 or residues 597-636 of SEQ ID NO: 2, in a pharmaceutically acceptable carrier.

FIG 1

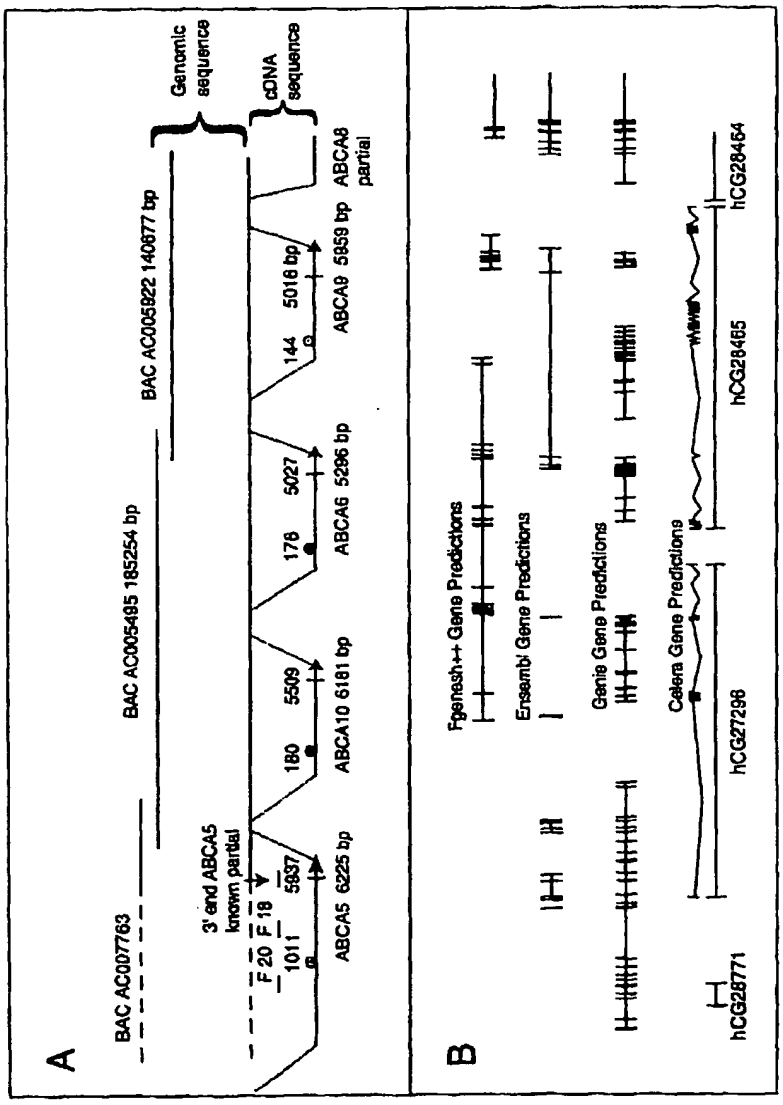
A
BAC AC007763
BAC AC005495 185254 bp
BAC AC005822 140677 bp
Genomic sequence
cDNA sequence
3' end ABCA5 known partial
F 20
F 18
1011
5937
180
5509
178
5027
144
5016 bp
ABCA5 6225 bp
ABCA10 6181 bp
ABCA6 5296 bp
ABCA9 5859 bp
ABCA8 partial
B
Fgenesh++ Gene Predictions
Ensembl Gene Predictions
Genie Gene Predictions
Celera Gene Predictions
hCG28771
hCG27296
hCG28465
hCG28464

FIG 2

```
                1                                                   50
huest698739  MVKKEISVRQ QIQALLYKNF LKKWRIKREF IEEWTITLFL GLYLCIFSEH
huest990006  MRKRKISVCQ QTWALLCKNF LKKWRMKRES LMEWLNSLLL LLCLYI.YPH
huest640918  MSKRRMSVGQ QTWALLCKNC LKKWRMKRQT LLEWLFSFLL VLFLYLFFSN
huest155051  MNMKQKSVYQ QTKALLCKNF LKKWRMKRES LLEWGLSILL GLCIALFSSS

                51                                                  100
huest698739  FRATR.FPEQ PPKVLGSVDQ FNDSGLVVAY TPVSNITQRI MNKMALASFM
huest990006  SHQVNDFSSL LTMDLGRVDT FNESRFSVVY TPVTNTTQQI MNKVASTPFL
huest640918  LHQVHDTPQM SSMDLGRVDS FNDTNYVIAF APESKTTQEI MNKVASAPFL
huest155051  MRNVQ.FPGM APQNLGRVDK FNSSSLMVVY TPISNLTQQI MNKTALAPLL

                101                                                 150
huest698739  KGRTVIGTPD EETMDIELPK KYHEMVGVIF SDTFSYRLKF NWGYRIPVIK
huest990006  AGKEVLGLPD EESIKEFTAN YPEEIVRVTF TNTYSYHLKF LLGHGMPAKK
huest640918  KGRTIMGWPD EKSMDELDLN YSIDAVRVIF TDTFSYHLKF SWGHRIPMMK
huest155051  KGTSVIGAPN KTHMDEILLE NLPYAMGIIF NETFSYKLIF FQGYNSPLWK

                151                                             200
huest698739  EHSEYTEHCW AMHGEIFCYL AKYWLKGFVA FQAAINAAII EVTTNHSVME
huest990006  EHKDHTAHCY ETNEDVYCEV SVFWKEGFVA LQAAINAAII EITTNHSVME
huest640918  EHRDHSAHCQ AVNEKMKCEG SEFWEKGFVA FQAAINAAII EIATNHSVME
huest155051  E..DFSAHCW DGYGEFSCTL TKYWNRGFVA LQTAINTAII EITTNHPVME

                201                                                 250
huest698739  ELTSVIGINM KIPPFISKGE IMNEWFHFTC LVSFSSFIYF ASLNVARERG
huest990006  ELMSVTGKNM KMHSFIGQSG VITDLYLFSC IISFSSFIYY ASVNVTRERK
huest640918  QLMSVTGVHM KILPFVAQGG VATDFFIFFC IISFSTFIYY VSVNVTQERQ
huest155051  ELMSVTAITM KTLPFITKNL LHNEMFILFF LLHFSPLVYF ISLNVTKERK

                251                                                 300
huest698739  KFKKLMTVMG LRESAFWLSW GLTYICFIFI MSIFMALVIT SIPIVFHTGF
huest990006  RMKALMTMMG LRDSAFWLSW GLLYAGFIFI MALFLALVIR STQFIILSGF
huest640918  YITSLMTMMG LRESAFWLSW GLMYAGFILI MATLMALIVK SAQIVVLTGF
huest155051  KSKNLMKMMG LQDSAFWLSW GLIYAGFIFI ISIFVTIIIT FTQIIVMTGF

                301                                                 350
huest698739  MVIFTLYSLY GLSLIALAFL MSVLIRKPML AGLAGFLFTV FWGCLGFTVL
huest990006  MVVFSLFLLY GLSLVALAFL MSILVKKSFL TGLVVFLLTV FWGCLGFTSL
huest640918  VMVFTLFLLY GLSLITLAFL MSVLIKKPFL TGLVVFLLIV FWGILGFPAL
huest155051  MVIFILFFLY GLSLVALVFL MSVLLKKAVL TNLVVFLLTL FWGCLGFTVF

                351                                                 400
huest698739  YRQLPLSLGW VLSLLSPFAF TAGMAQITHL DNYLSGVIFP DPSGDSYKMI
huest990006  YRHLPASLEW ILSLLSPFAF MLGMAQLLHL DYDLNSNAFP HPSDGSNLIV
huest640918  YTRLPAFLEW TLCLLSPFAF TVGMAQLIHL DYDVNSNAHL DSSQNPYLII
huest155051  YEQLPSSLEW ILNICSPFAF TTGMIQIIKL DYNLNGVIFP DPSGDSYTMI
```

FIG 2

```
                401                                                 450
huest698739  ATFFILAFDT LFYLIFTLYF ERVLPDKDGH GDSPLFFLKS SFWSKHQNTH
huest990006  ATNFMLAFDT CLYLALAIYF EKILPNEYGH RRPPLFFLKS SFWSQTQKTD
huest640918  ATLFMLVFDT LLYLVLTLYF DKILPAEYGH RCSPLFFLKS CFWFQHGRAN
huest155051  ATFSMLLLDG LIYLLLALYF DKILPYGDER HYSPLFFLNS SSCFQHQRTN

                451                                                 500
huest698739  HEIFENEINP EHSSDDSFEP VSPEFHGKEA IRIRNVIKEY NGKTGKVEAL
huest990006  HVALEDEMDA DPSFHDSFEQ APPEFQGKEA IRIRNVTKEY KGKPDKIEAL
huest640918  HVVLENETDS DPTPNDCFEP VSPEFCGKEA IRIKNLKKEY AGKCERVEAL
huest155051  AKVIEKEIDA EHPSDDYFEP VAPEFQGKEA IRIRNVKKEY KGKSGKVEAL

                501                                                 550
huest698739  QGIFFDIYEG QITAILGHNG AGKSTLLNIL SGLSVSTEGS ATIYNTQLSE
huest990006  KDLVFDIYEG QITAILGHSG AGKSTLLNIL SGLSVPTKGS VTIYNNKLSE
huest640918  KGVVFDIYEG QITALLGHSG AGKTTLLNIL SGLSVPTSGS VTVYNHTLSR
huest155051  KGLLFDIYEG QITAILGHSG AGKSSLLNIL NGLSVPTEGS VTIYNKNLSE

                551                                                 600
huest698739  ITDMEEIRKN IGFCPQFNFQ FDFLTVRENL RVFAKIKGIQ PKEVEQEVKR
huest990006  MADLENLSKL TGVCPQSNVQ FDFLTVRENL RLFAKIKGIL PQEVDKEIQR
huest640918  MADIENISKF TGFCPQSNVQ FGFLTVKENL RLFAKIKGIL PHEVEKEVQR
huest155051  MQDLEEIRKI TGVCPQFNVQ FDILTVKENL SLFAKIKGIH LKEVEQEVQR

                601                                                 650
huest698739  IIMELDMQSI QDIIAKKLSG GQKRKLTLGI AILGDPQVLL LDEPTAGLDP
huest990006  VLLELEMKNI QDVLAQNLSG GQKRKLTFGI AILGDPQIFL LDEPTAGLDP
huest640918  VVQELEMENI QDILAQNLSG GQNRKLTFGI AILGDPQVLL LDEPTAGLDP
huest155051  ILLELDMQNI QDNLAKHLSE GQKRKLTFGI TILGDPQILL LDEPTTGMVP

                651                                                 700
huest698739  FSRHRVWSLL KEHKVDRLIL FSTQFMDEAD ILADRKVFLS NGKLKCAGSS
huest990006  FSRHQVWNLL KERKTDRVIL FSTQFMDEAD ILADRKVFLS QGKLKCAGSS
huest640918  LSRHRIWNLL KEGKSDRVIL FSTQFIDEAD ILADRKVFIS NGKLKCAGSS
huest155051  FSRDQVWSLL RERRADHVIL FSTQSMDEAD ILADRKVIMS NGRLKCAGSS

                701                                                 750
huest698739  LFLKRKWGIG YHLSLHRNEM CDTEKITSLI KQHIPDAKLT TESEEKLVYS
huest990006  LFLKKKWGIG YHLSLQLNEI CVEENITSLV KQHIPDAKLS AKSEGKLIYT
huest640918  LFLKKKWGIG YHLSLHLNER CDPESITSLV KQHISDAKLT AQSEEKLVYI
huest155051  MFLKRRWGLG YHLSLHRNEI CNPEQITSFI THHIPDAKLK TENKEKLVYT

                751                                                 800 *
huest698739  LPLEKTNKFP DLYSDLDKCS DQGIRNYAVS VTSLNEVFLN LEGKSAIDEP
huest990006  LPLERTNKFP ELYKDLDSYP DLGIENYGVS MTTLNEVFLK LEGKSTINES
huest640918  LPLERTNKFP ELYRDLDRCS NQGIEDYGVS ITTLNEVFLK LEGKSTIDES
huest155051  LPLERTNTFP DLFSDLDKCS DQGVTGYDIS MSTLNEVFMK LEGQSTIEQD
```

FIG 2

```
            801                                                  850
huest698739 GIKMYFDIGK QEKIHVTRNT GDESEMEQVL CSLPETRKAV SSAALWRRQI
huest990006 DIAILGEVQA EK....ADDT ERLVEMEQVL SSLNKMRKTI GGVALWRQQI
huest640918 DIGIWGQLQT DG....AKDI GSLVELEQVL SSFHETRKTI SGVALWRQQV
huest155051 .......... FEQVEMIRDS ESLNEMELAH SSFSEMQTAV SDMGLWRMQV

            851                                                  900
huest698739 YAVATLRFLK LRRERRALLC LLLVLGIAFI PIILEKIMYK VTRETHCWEF
huest990006 CAIARVRLLK LKHERKALLA LLLILMAGFC PLLVEYTMVK IYQNSYTWEL
huest640918 CAIAKVRFLK LKKERKSLWT ILLLFGISFI PQLLEHLFYE SYQKSYPWEL
huest155051 FAMARLRFLK LKRQTKVLLT LLLVFGIAIF PLIVENIMYA MLNEKIDWEF

            901                                                  950
huest698739 SPSMYFLSLE QIPKTPLTSL LIVNNTGSNI EDLVHSLKCQ DIVLEIDDFR
huest990006 SPHLYFLAPG QQPHDPLTQL LIINKTGASI DDFIQSVEHQ NIALEVDAFG
huest640918 SPNTYFLSPG QQPQDPLTHL LVINKTGSTI DNFLHSLRRQ NIAIEVDAFG
huest155051 KNELYFLSPG QLPQEPRTSL LIINNTESNI EDFIKSLKHQ NILLEVDDFE

            951                                                 1000
huest698739 NRNGSDDPSY NGAIIVSGDQ KDYRFSVACN TKKLNCFPVL MGIVSNALMG
huest990006 TRNGTDDPSY NGAITVCCNE KNYSFSLACN AKRLNCFPVL MDIVSNGLLG
huest640918 TRNGTDDPSY NGAIIVSGDE KDHRFSIACN TKRLNCFPVL LDVISNGLLG
huest155051 NRNGTDGLSY NGAIIVSGKQ KDYRFSVVCN TKRLHCFPIL MNIISNGLLQ

            1001                                                1050
huest698739 IFNFTELIQM ESTSFSRDDI VLDLGFIDGS IFLLLITNCV SPFIGMSSIS
huest990006 MVKPSVHIRT ERSTFLENGQ DNPIGFLAYI MFWLVLTSSC PPYIAMSSID
huest640918 IFNSSEHIQT DRSTFFEEHM DYEYGYRSNT FFWIPMAASF TPYIAMSSIG
huest155051 MFNHTQHIRI ESSPFPLSHI GLWTGLPDGS FFLFLVLCSI SPYITMGSIS

            1051                                                1100
huest698739 DYKKNVQSQL WISGLWPSAY WCGQALVDIP LYFLILFSIH LI.YYFIFLG
huest990006 DYKNRARSQL RISGLSPSAY WFGQALVDVS LYFLVFVFIY LMSYISNFED
huest640918 DYKKKAHSQL RISGLYPSAY WFGQALVDVS LYFLILLLMQ IMDYIFSPEE
huest155051 DYKKNAKSQL WISRPLHFCL LVWAGTMDVS FFILILLLMY LIFYIENMQY

            1101                                                1150
huest698739 FQLSWELMFV LVVCIIGCAV SLIFLTYVLS FIFRKWRKNN GFWSFGFFII
huest990006 MLLTIIH.II QIPCAVGYSF SLIFMTYVIS FIFRKGRKNS GIWSFCFYVV
huest640918 IIFIIQNLLI QILCSIGYVS SLVFLTYVIS FIFRNGRKNS GIWSFFFLIV
huest155051 LLITSQIVFA LVIVTPGYAA SLVFFIYMIS FIFRKRRKNS GLWSFYFFFA

            1151                                                1200
huest698739 LICVSYIMVS TQYEKLNLIL CMIFIPSFTL LGYVMLLIQL DFMRNLDSLD
huest990006 TVFSVAGFAF SIFESDIPFI FTFLIPPATM IGCLFLSSHL LF.SSL..FS
huest640918 VIFSIVATDL NEYGFLGLFF GTMLIPPFTL IGSLFIFSEI SP.DSMDYLG
huest155051 STIMFSITLI NHFDLSILIT TMVLVPSYTL LGFK.TFLEV RDQEHYREFP
```

FIG 2

```
                1201                                                1250
huest698739     NRINEVNKTI LLTTLIPYLQ SVIFLFVIRC LEMKYGNEIM NKDPVFR.IS
huest990006     EERMDVQ..P FLVFLIPFLH FIIFLFTLRC LEWKFGKKSM RKDPFFR.IS
huest640918     ASESEI...V YLALLIPYLH FLIFLFILRC LEMNCRKKLM RKDPVFR.IS
huest155051     EANFELSATD FLVCFIPYFQ TLLFVFVLRC MELKCGKKRM RKDPVFR.IS
 huest90625     ---------- ------PYLQ CVLWIFLLQY YEKKYGGRSI RKDPFFRNLS

                1251                                                1300
huest698739     PRSRETH.PN PEEPEEEDED VQAERVQAAN ALTAPNLEEE PVITASCLHK
huest990006     PRSSDVC.QN PEEPEGEDED VQMERVRTAN ALNSTNFDEK PVIIASCLRK
huest640918     PRSNAIF.PN PEEPEGEEED IQMERMRTVN AMAVRDFDET PVIIASCLRK
huest155051     PQSRDAK.PN PEEPIDEDED IQTERIRTAT ALTTSILDEK PVIIASCLHK
 huest90625     TKSKNRKLPE PPDNEDEDED VKAERLKVKE LMGCQCCEEK PSIMVSNLHK

                1301                                                1350
huest1133530    ---------- ---------- ---------- --GLLGHNGA GKSTSIKMIT
 huest698739    EYYETKKSCF STRKKKIAIR NVSFCVKKGE VLGLLGHNGA GKSTSIKMIT
 huest990006    EYAGKRKGCF SKRKNKIATR NVSFCVRKGE VLGLLGHNGA GKSTSIKVIT
 huest640918    EYAGKKKNCF SKRKKKIATR NVSFCVKKGE VIGLLGHNGA GKSTTIKMIT
 huest155051    EYAGQKKSCF SKRKKKIAAR NISFCVQEGE ILGLLGPNGA GKSSSIRMIS
  huest90625    EYDDKKDFLL SRKVKKVATK YISFCVKKGE ILGLLGPNGA GKSTIINILV

                1351                                                1400
huest1133530    GCTVPTAGVV VLQGNRASVR QQRDNSLKXL GYCPQENSLC PKLTMKEHLE
 huest698739    GCTKPTAGVV VLQGSRASVR QQHDNSLKFL GYCPQENSLW PKLTMKEHLE
 huest990006    GDTKPTAGQV LLKGSGGG.. ....DALEFL GYCPQENALW PNLTVRQHLE
 huest640918    GDTKPTAGQV ILKGSGGG.. ....EPLGFL GYCPQENALW PNLTVRQHLE
 huest155051    GITKPTAGEV ELKG...... ..CSSVLGHL GYCPQENVLW PMLTLREHLE
  huest90625    GDIEPTSGQV FL.GDYSSET SEDDDSLKCM GYCPQINPLW PDTTLQEHFE

                1401                                                1450
huest1133530    LYAAVKGLGK .DAALSISXL VEALKLQEQL KAPVKTLSEG IKRKLCFVLS
 huest698739    LYAAVKGLGK EDAALSISRL VEALKLQEQL KAPVKTLSEG IKRKLCFVLS
 huest990006    VYAAVKGLRK GDAEVAITRL VDALKLQDQL KSPVKTLSEG IKRKLCFVLS
 huest640918    VYAAVKGLRK GDAMIAITRL VDALKLQDQL KAPVKTLSEG IKRKLCFVLS
 huest155051    VYAAVKGLRK ADARLAIARL VSAFKLHEQL NVPVQKLTAG ITRKLCFVLS
  huest90625    IYGAVKGMSA SDMKEVISRI THALDLKEHL QKTVKKLPAG IKRKLCFALS

                1451                                                1500
huest1133530    ILGNPSVVLL DELFTGMDPE GQQQMWQILQ ATIKNQERGA LLTTHYMSEA
 huest698739    ILGNPSVVLL DEPFTGMDPE GQQQMWQILQ ATVKNKERGT LLTTHYMSEA
 huest990006    ILGNPSVVLL DEPSTGMDPE GQQQMWQAIR ATFRNTERGA LLTTHYMAEA
 huest640918    ILGNPSVVLL DEPSTGMDPE GQQQMWQVIR ATFRNTERGA LLTTHYMAEA
 huest155051    LLGNSPVLLL DEPSTGITPQ GSSKCWQAIQ AVVKNTERGV LLTTHNLAEA
  huest90625    MLGNPQITLL DEPSTGMDPK AKQHMWRAIR TAFKNRKRAA ILTTHYMEEA
```

FIG 2

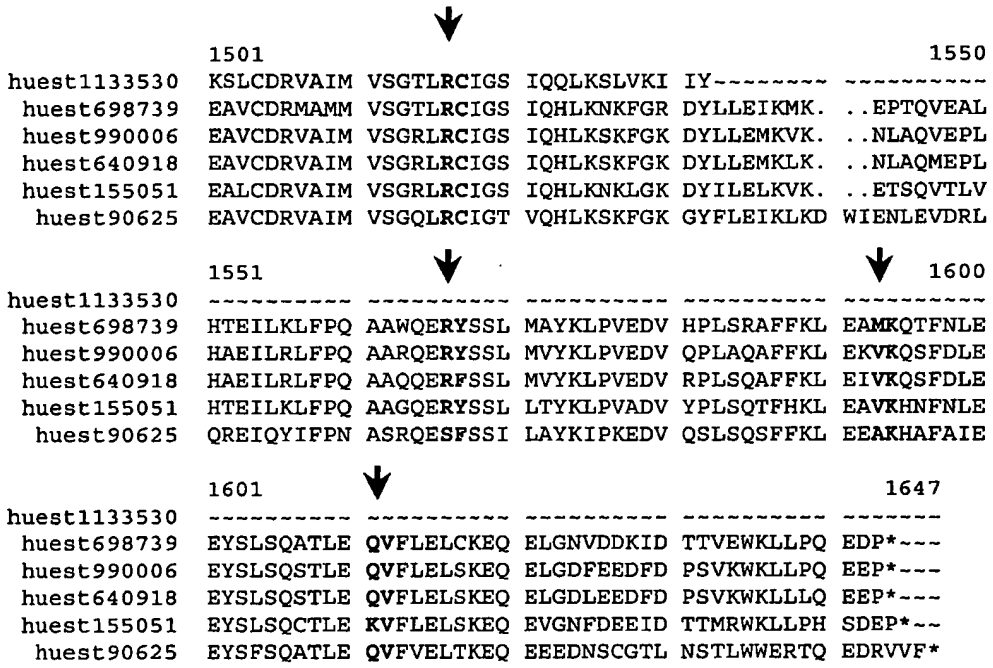
1501
1550
huest1133530 KSLCDRVAIM VSGTLRCIGS IQQLKSLVKI IY~~~~~~~~ ~~~~~~~~~~
huest698739 EAVCDRMAMM VSGTLRCIGS IQHLKNKFGR DYLLEIKMK. ..EPTQVEAL
huest990006 EAVCDRVAIM VSGRLRCIGS IQHLKSKFGK DYLLEMKVK. ..NLAQVEPL
huest640918 EAVCDRVAIM VSGRLRCIGS IQHLKSKFGK DYLLEMKLK. ..NLAQMEPL
huest155051 EALCDRVAIM VSGRLRCIGS IQHLKNKLGK DYILELKVK. ..ETSQVTLV
huest90625 EAVCDRVAIM VSGQLRCIGT VQHLKSKFGK GYFLEIKLKD WIENLEVDRL
1551
1600
huest1133530 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~
huest698739 HTEILKLFPQ AAWQERYSSL MAYKLPVEDV HPLSRAFFKL EAMKQTFNLE
huest990006 HAEILRLFPQ AARQERYSSL MVYKLPVEDV QPLAQAFFKL EKVKQSFDLE
huest640918 HAEILRLFPQ AAQQERFSSL MVYKLPVEDV RPLSQAFFKL EIVKQSFDLE
huest155051 HTEILKLFPQ AAGQERYSSL LTYKLPVADV YPLSQTFHKL EAVKHNFNLE
huest90625 QREIQYIFPN ASRQESFSSI LAYKIPKEDV QSLSQSFFKL EEAKHAFAIE
1601
1647
huest1133530 ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~~~~ ~~~~~~~
huest698739 EYSLSQATLE QVFLELCKEQ ELGNVDDKID TTVEWKLLPQ EDP*~~~
huest990006 EYSLSQSTLE QVFLELSKEQ ELGDFEEDFD PSVKWKLLPQ EEP*~~~
huest640918 EYSLSQSTLE QVFLELSKEQ ELGDLEEDFD PSVKWKLLLQ EEP*~~~
huest155051 EYSLSQCTLE KVFLELSKEQ EVGNFDEEID TTMRWKLLPH SDEP*~~
huest90625 EYSFSQATLE QVFVELTKEQ EEEDNSCGTL NSTLWWERTQ EDRVVF*

FIG 3A

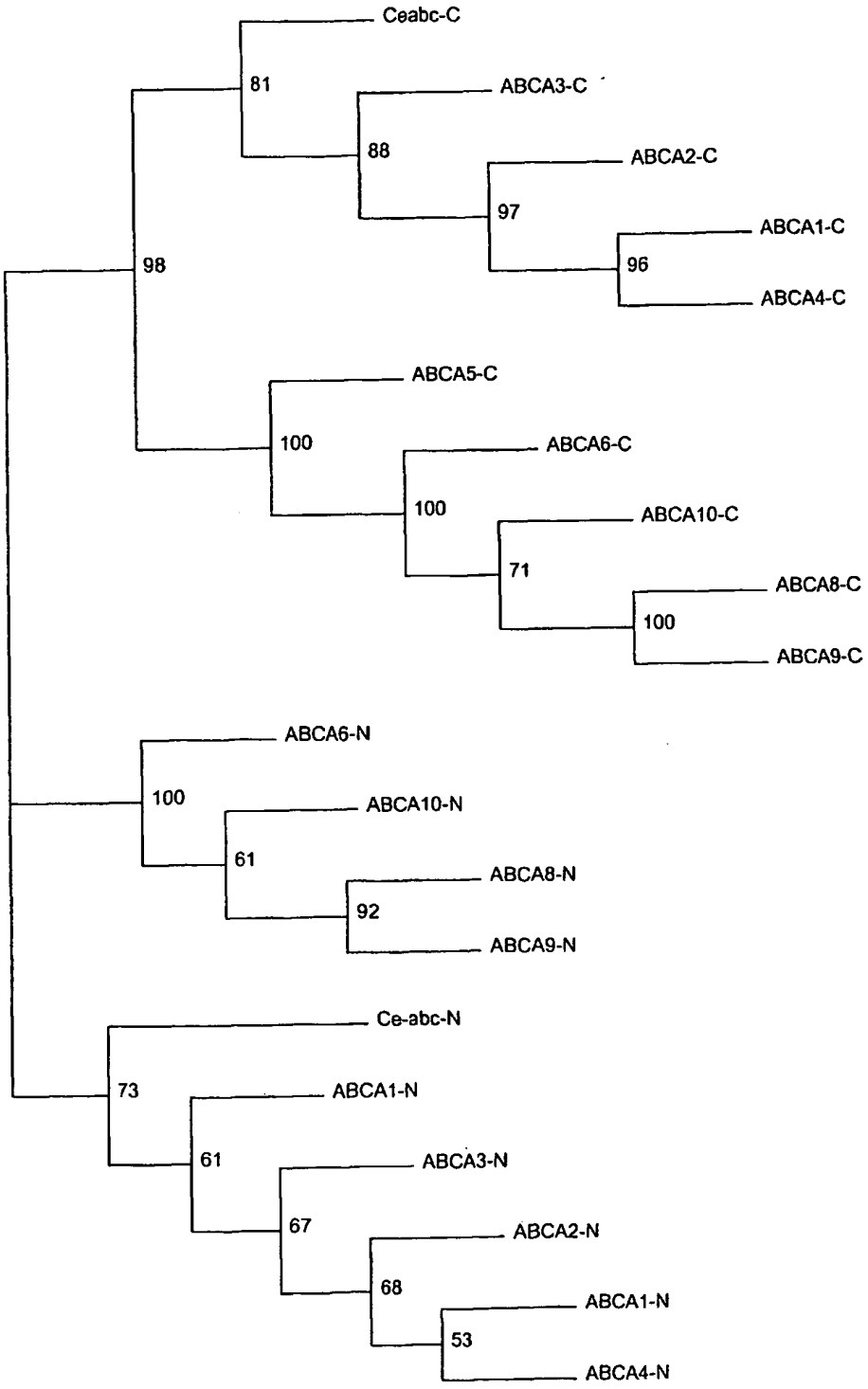
Ceabc-C
81
ABCA3-C
88
ABCA2-C
97
ABCA1-C
96
ABCA4-C
98
ABCA5-C
100
ABCA6-C
100
ABCA10-C
71
ABCA8-C
100
ABCA9-C
ABCA6-N
100
ABCA10-N
61
ABCA8-N
92
ABCA9-N
Ce-abc-N
73
ABCA1-N
61
ABCA3-N
67
ABCA2-N
68
ABCA1-N
53
ABCA4-N

FIG 3B

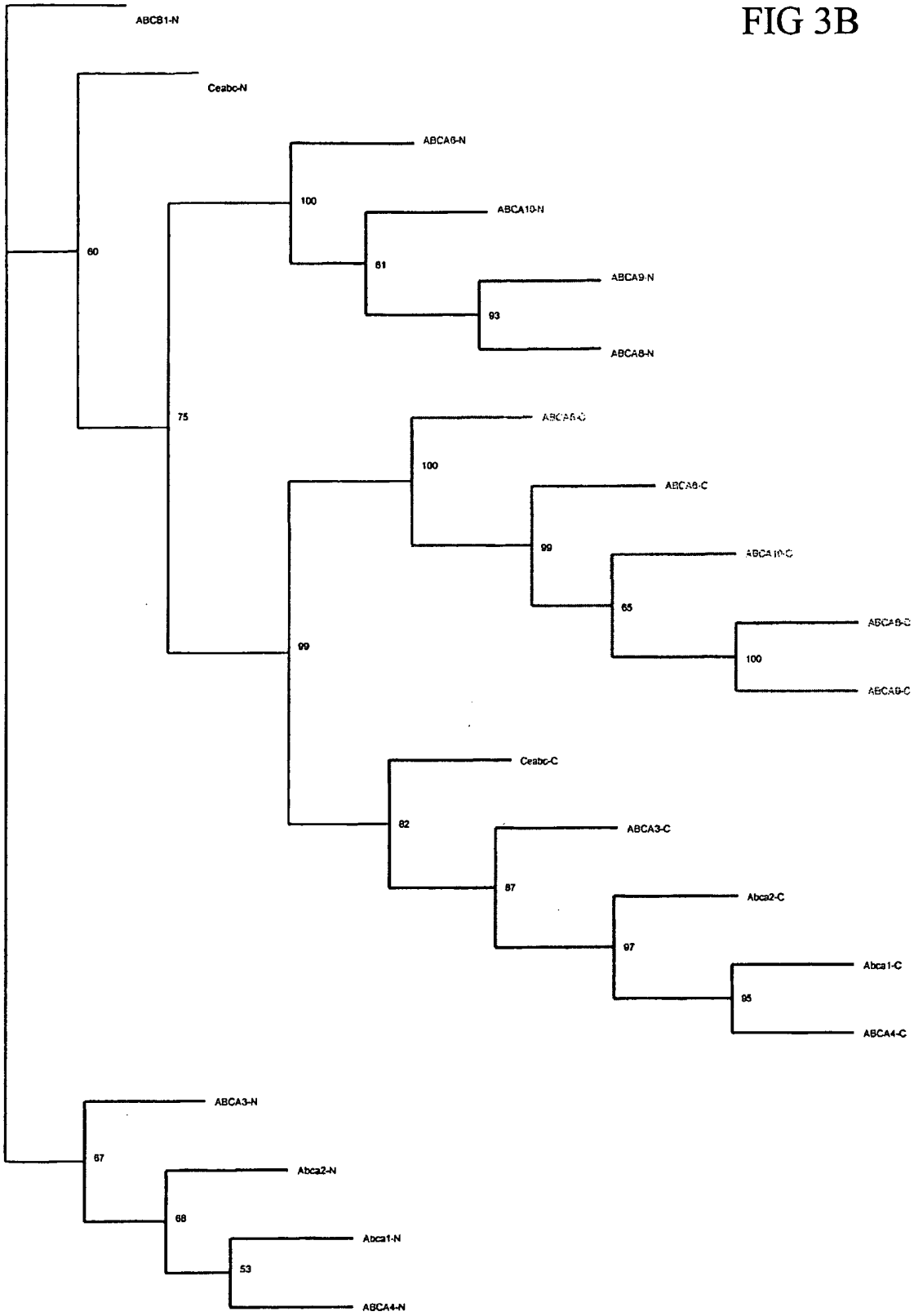
ABCB1-N
Ceabc-N
ABCA6-N
ABCA10-N
ABCA9-N
ABCA8-N
ABCA5-C
ABCA6-C
ABCA10-C
ABCA9-C
ABCA8-C
Ceabc-C
ABCA3-C
Abca2-C
Abca1-C
ABCA4-C
ABCA3-N
Abca2-N
Abca1-N
ABCA4-N
100
60
61
93
75
100
99
65
100
99
82
87
97
95
67
68
53

FIG 4A

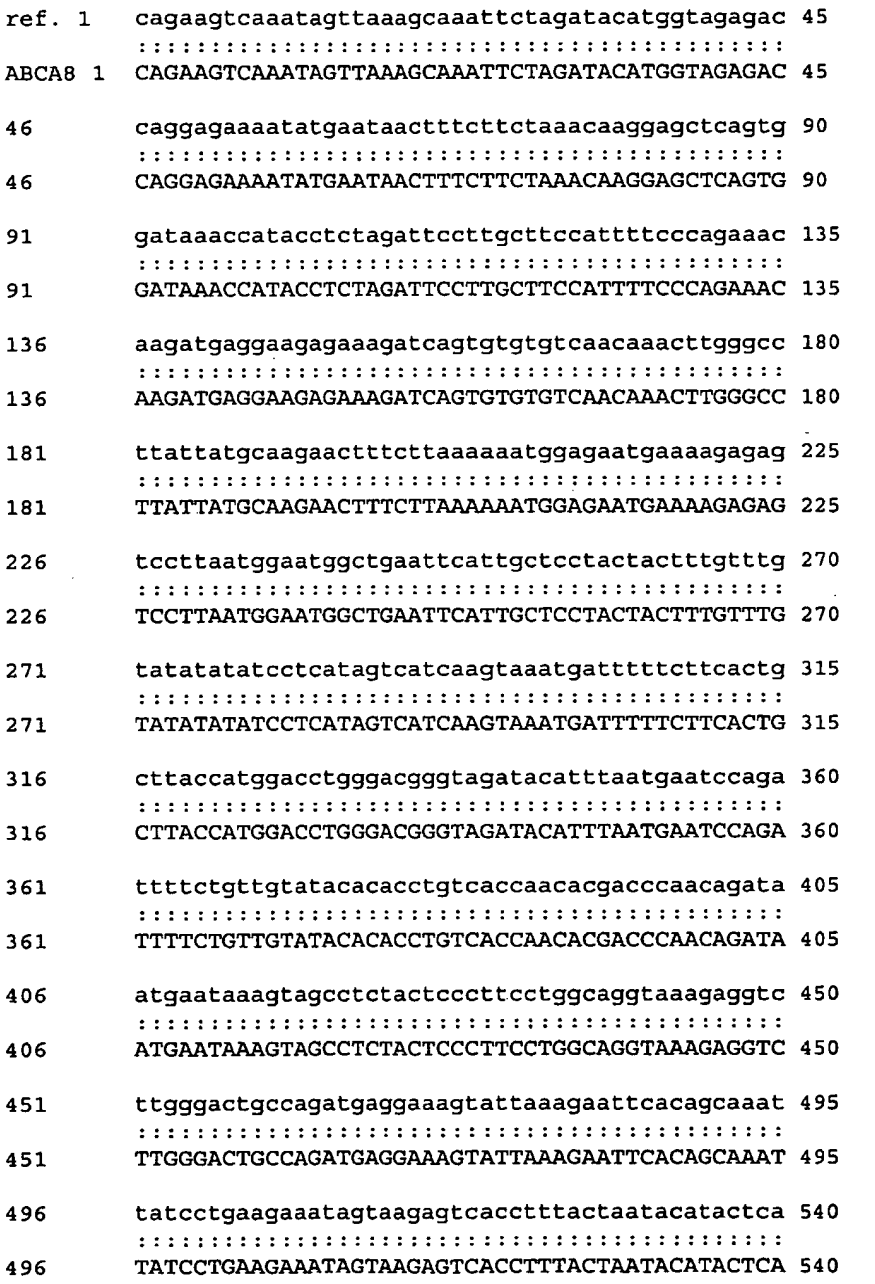
ref. 1 cagaagtcaaatagttaaagcaaattctagatacatggtagagac 45
ABCA8 1 CAGAAGTCAAATAGTTAAAGCAAATTCTAGATACATGGTAGAGAC 45
46 caggagaaaatatgaataactttcttctaaacaaggagctcagtg 90
46 CAGGAGAAAATATGAATAACTTTCTTCTAAACAAGGAGCTCAGTG 90
91 gataaaccatacctctagattccttgcttccattttcccagaaac 135
91 GATAAACCATACCTCTAGATTCCTTGCTTCCATTTTCCCAGAAAC 135
136 aagatgaggaagagaaagatcagtgtgtgtcaacaaacttgggcc 180
136 AAGATGAGGAAGAGAAAGATCAGTGTGTGTCAACAAACTTGGGCC 180
181 ttattatgcaagaactttcttaaaaaatggagaatgaaaagagag 225
181 TTATTATGCAAGAACTTTCTTAAAAAATGGAGAATGAAAAGAGAG 225
226 tccttaatggaatggctgaattcattgctcctactactttgtttg 270
226 TCCTTAATGGAATGGCTGAATTCATTGCTCCTACTACTTTGTTTG 270
271 tatatatatcctcatagtcatcaagtaaatgatttttcttcactg 315
271 TATATATATCCTCATAGTCATCAAGTAAATGATTTTTCTTCACTG 315
316 cttaccatggacctgggacgggtagatacatttaatgaatccaga 360
316 CTTACCATGGACCTGGGACGGGTAGATACATTTAATGAATCCAGA 360
361 ttttctgttgtatacacacctgtcaccaacacgacccaacagata 405
361 TTTTCTGTTGTATACACACCTGTCACCAACACGACCCAACAGATA 405
406 atgaataaagtagcctctactcccttcctggcaggtaaagaggtc 450
406 ATGAATAAAGTAGCCTCTACTCCCTTCCTGGCAGGTAAAGAGGTC 450
451 ttgggactgccagatgaggaaagtattaaagaattcacagcaaat 495
451 TTGGGACTGCCAGATGAGGAAAGTATTAAAGAATTCACAGCAAAT 495
496 tatcctgaagaaatagtaagagtcacctttactaatacatactca 540
496 TATCCTGAAGAAATAGTAAGAGTCACCTTTACTAATACATACTCA 540

FIG 4A

```
541     tatcatttgaagttcttgctaggacatggaatgccagcaaagaag 585
        :::::::::::::::::::::::::::::::::::::::::::::
541     TATCATTTGAAGTTCTTGCTAGGACATGGAATGCCAGCAAAGAAG 585

586     gagcacaaggaccatacagctcattgttatgaaacaaatgaagat 630
        :::::::::::::::::::::::::::::::::::::::::::::
586     GAGCACAAGGACCATACAGCTCATTGTTATGAAACAAATGAAGAT 630

631     gtttactgtgaagtttcagtattttggaaggaaggttttgtggct 675
        :::::::::::::::::::::::::::::::::::::::::::::
631     GTTTACTGTGAAGTTTCAGTATTTTGGAAGGAAGGTTTTGTGGCT 675

676     cttcaagctgccattaatgctgctattatagaaatcacaacaaat 720
        :::::::::::::::::::::::::::::::::::::::::::::
676     CTTCAAGCTGCCATTAATGCTGCTATTATAGAAATCACAACAAAT 720

721     cactcagtgatggaggagctgatgtcagttactggaaaaaatatg 765
        :::::::::::::::::::::::::::::::::::::::::::::
721     CACTCAGTGATGGAGGAGCTGATGTCAGTTACTGGAAAAAATATG 765

766     aagatgcattccttcattggtcaatcaggagttataactgatttg 810
        :::::::::::::::::::::::::::::::::::::::::::::
766     AAGATGCATTCCTTCATTGGTCAATCAGGAGTTATAACTGATTTG 810

811     tacctttttttcctgcattatttcattttcctcattcatttactat 855
        :::::::::::::::::::::::::::::::::::::::::::::
811     TACCTTTTTTCCTGCATTATTTCATTTTCCTCATTCATTTACTAT 855

856     gcatctgttaatgtcacaagagagaggaaaaggatgaaggccttg 900
        :::::::::::::::::::::::::::::::::::::::::::::
856     GCATCTGTTAATGTCACAAGAGAGAGGAAAAGGATGAAGGCCTTG 900

901     atgacaatgatgggtcttcgggattcagcgttctggctctcctgg 945
        :::::::::::::::::::::::::::::::::::::::::::::
901     ATGACAATGATGGGTCTTCGGGATTCAGCGTTCTGGCTCTCCTGG 945

946     ggtttgctctatgctggtttcatcttcattatggccctttttcttg 990
        :::::::::::::::::::::::::::::::::::::::::::::
946     GGTTTGCTCTATGCTGGTTTCATCTTCATTATGGCCCTTTTCTTG 990

991     gcacttgttataagatctacccagtttatcattttgtctggcttc 1035
        :::::::::::::::::::::::::::::::::::::::::::::
991     GCACTTGTTATAAGATCTACCCAGTTTATCATTTTGTCTGGCTTC 1035

1036    atggtagtcttcagcctctttctcctgtatggattatctttggta 1080
        :::::::::::::::::::::::::::::::::::::::::::::
1036    ATGGTAGTCTTCAGCCTCTTTCTCCTGTATGGATTATCTTTGGTA 1080
```

FIG 4A

```
1081    gctttggctttcttaatgagcatcttggtaaagaaatctttcctc 1125
        :::::::::::::::::::::::::::::::::::::::::::
1081    GCTTTGGCTTTCTTAATGAGCATCTTGGTAAAGAAATCTTTCCTC 1125

1126    accggcctggtcgtgttcctcctcactgtcttttgggggtgtctg 1170
        :::::::::::::::::::::::::::::::::::::::::::
1126    ACCGGCCTGGTCGTGTTCCTCCTCACTGTCTTTTGGGGGTGTCTG 1170

1171    gggttcacatcactgtacagacaccttcctgcatccttggagtgg 1215
        :::::::::::::::::::::::::::::::::::::::::::
1171    GGGTTCACATCACTGTACAGACACCTTCCTGCATCCTTGGAGTGG 1215

1216    attttaagcttgcttagtccctttgccttcatgcttggaatggcc 1260
        :::::::::::::::::::::::::::::::::::::::::::
1216    ATTTTAAGCTTGCTTAGTCCCTTTGCCTTCATGCTTGGAATGGCC 1260

1261    cagcttttacacttggactatgatttgaattctaatgcatttcct 1305
        :::::::::::::::::::::::::::::::::::::::::::
1261    CAGCTTTTACACTTGGACTATGATTTGAATTCTAATGCATTTCCT 1305

1306    catccatcggacggctcaaatctcattgtagcaacaaatttcatg 1350
        :::::::::::::::::::::::::::::::::::::::::::
1306    CATCCATCGGACGGCTCAAATCTCATTGTAGCAACAAATTTCATG 1350

1351    ttggcatttgacacttgcctctatctggcattggcgatttacttt 1395
        :::::::::::::::::::::::::::::::::::::::::::
1351    TTGGCATTTGACACTTGCCTCTATCTGGCATTGGCGATTTACTTT 1395

1396    gaaaaaattttgccaaatgaatatggacatcgacgtccacctttg 1440
        :::::::::::::::::::::::::::::::::::::::::::
1396    GAAAAAATTTTGCCAAATGAATATGGACATCGACGTCCACCTTTG 1440

1441    tttttcctgaagtcctcattttggtctcaaacacaaaagactgat 1485
        :::::::::::::::::::::::::::::::::::::::::::
1441    TTTTTCCTGAAGTCCTCATTTTGGTCTCAAACACAAAAGACTGAT 1485

1486    cacgtggcccttgaagatgaaatggatgccgatccttcatttcat 1530
        :::::::::::::::::::::::::::::::::::::::::::
1486    CACGTGGCCCTTGAAGATGAAATGGATGCCGATCCTTCATTTCAT 1530

1531    gactcttttgaacaagcgcctccagaattccaagggaaagaagcc 1575
        :::::::::::::::::::::::::::::::::::::::::::
1531    GACTCTTTTGAACAAGCGCCTCCAGAATTCCAAGGGAAAGAAGCC 1575

1576    atcagaatcagaaatgttacaaaagaatataaaggaaagcctgat 1620
        :::::::::::::::::::::::::::::::::::::::::::
1576    ATCAGAATCAGAAATGTTACAAAAGAATATAAAGGAAAGCCTGAT 1620
```

FIG 4A

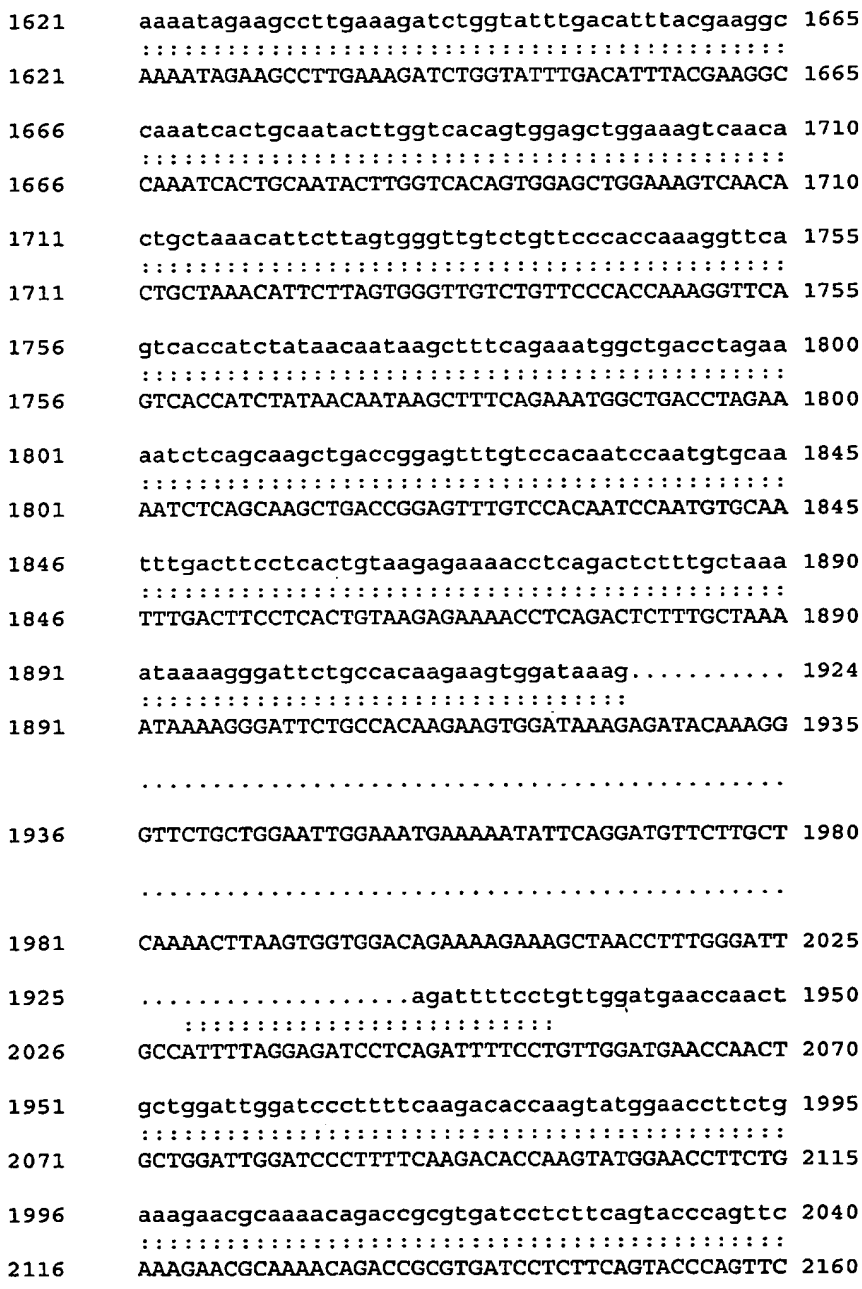
1621 aaaatagaagccttgaaagatctggtatttgacatttacgaaggc 1665
1621 AAAATAGAAGCCTTGAAAGATCTGGTATTTGACATTTACGAAGGC 1665
1666 caaatcactgcaatacttggtcacagtggagctggaaagtcaaca 1710
1666 CAAATCACTGCAATACTTGGTCACAGTGGAGCTGGAAAGTCAACA 1710
1711 ctgctaaacattcttagtgggttgtctgttcccaccaaaggttca 1755
1711 CTGCTAAACATTCTTAGTGGGTTGTCTGTTCCCACCAAAGGTTCA 1755
1756 gtcaccatctataacaataagctttcagaaatggctgacctagaa 1800
1756 GTCACCATCTATAACAATAAGCTTTCAGAAATGGCTGACCTAGAA 1800
1801 aatctcagcaagctgaccggagtttgtccacaatccaatgtgcaa 1845
1801 AATCTCAGCAAGCTGACCGGAGTTTGTCCACAATCCAATGTGCAA 1845
1846 tttgacttcctcactgtaagagaaaacctcagactctttgctaaa 1890
1846 TTTGACTTCCTCACTGTAAGAGAAAACCTCAGACTCTTTGCTAAA 1890
1891 ataaaagggattctgccacaagaagtggataaag..... 1924
1891 ATAAAAGGGATTCTGCCACAAGAAGTGGATAAAGAGATACAAAGG 1935
.....
1936 GTTCTGCTGGAATTGGAAATGAAAAATATTCAGGATGTTCTTGCT 1980
.....
1981 CAAAACTTAAGTGGTGGACAGAAAAGAAAGCTAACCTTTGGGATT 2025
1925 .....agattttcctgttggatgaaccaact 1950
2026 GCCATTTTAGGAGATCCTCAGATTTTCCTGTTGGATGAACCAACT 2070
1951 gctggattggatccctttttcaagacaccaagtatggaaccttctg 1995
2071 GCTGGATTGGATCCCTTTTCAAGACACCAAGTATGGAACCTTCTG 2115
1996 aaagaacgcaaaacagaccgcgtgatcctcttcagtacccagttc 2040
2116 AAAGAACGCAAAACAGACCGCGTGATCCTCTTCAGTACCCAGTTC 2160

FIG 4A

```
2041    atggatgaggccgacatcctggcggacaggaaagtatttctctcc 2085
        :::::::::::::::::::::::::::::::::::::::::::::
2161    ATGGATGAGGCCGACATCCTGGCGGACAGGAAAGTATTTCTCTCC 2205

2086    caagggaagctaaagtgcgcgggctcttctttgtttctaaagaag 2130
        :::::::::::::::::::::::::::::::::::::::::::::
2206    CAAGGGAAGCTAAAGTGCGCGGGCTCTTCTTTGTTTCTAAAGAAG 2250

2131    aaatgggggattggatatcacttaagcttgcagttaaatgaaata 2175
        :::::::::::::::::::::::::::::::::::::::::::::
2251    AAATGGGGGATTGGATATCACTTAAGCTTGCAGTTAAATGAAATA 2295

2176    tgtgttgaggaaaacataacatcacttgttaaacagcacatccct 2220
        :::::::::::::::::::::::::::::::::::::::::::::
2296    TGTGTTGAGGAAAACATAACATCACTTGTTAAACAGCACATCCCT 2340

2221    gatgccaaattatcagccaaaagcgaaggaaaacttatttataca 2265
        :::::::::::::::::::::::::::::::::::::::::::::
2341    GATGCCAAATTATCAGCCAAAAGCGAAGGAAAACTTATTTATACA 2385

2266    ttacccttagaaagaacaaataaatttccagaactttacaaggat 2310
        :::::::::::::::::::::::::::::::::::::::::::::
2386    TTACCCTTAGAAAGAACAAATAAATTTCCAGAACTTTACAAGGAT 2430

2311    cttgatagctatcctgacctaggaattgagaattatggtgtttcc 2355
        :::::::::::::::::::::::::::::::::::::::::::::
2431    CTTGATAGCTATCCTGACCTAGGAATTGAGAATTATGGTGTTTCC 2475

2356    atgacaactttgaatgaagtattcctgaagctagaaggaaaatct 2400
        :::::::::::::::::::::::::::::::::::::::::::::
2476    ATGACAACTTTGAATGAAGTATTCCTGAAGCTAGAAGGAAAATCT 2520

2401    acaattaatgaatcggacattgctattttgggagaagtacaagcg 2445
        :::::::::::::::::::::::::::::::::::::::::::::
2521    ACAATTAATGAATCGGACATTGCTATTTTGGGAGAAGTACAAGCG 2565

2446    gaaaaagctgacgacactgaaaggcttgttgagatggaacaagtc 2490
        :::::::::::::::::::::::::::::::::::::::::::::
2566    GAAAAAGCTGACGACACTGAAAGGCTTGTTGAGATGGAACAAGTC 2610

2491    ctctcttcacttaacaagatgagaaagacaataggtggtgtggct 2535
        :::::::::::::::::::::::::::::::::::::::::::::
2611    CTCTCTTCACTTAACAAGATGAGAAAGACAATAGGTGGTGTGGCT 2655

2536    ctctggcgacagcaaatctgcgcaattgcaagggttcgcttgtta 2580
        :::::::::::::::::::::::::::::::::::::::::::::
2656    CTCTGGCGACAGCAAATCTGCGCAATTGCAAGGGTTCGCTTGTTA 2700
```

FIG 4A

```
2581    aagttaaagcatgaaagaaaagctcttttagcactgctattaatt 2625
        ::::::::::::::::::::::::::::::::::::::::::::
2701    AAGTTAAAGCATGAAAGAAAAGCTCTTTTAGCACTGCTATTAATT 2745

2626    ctaatggctggattttgccctcttcttgtggagtataccatggtg 2670
        ::::::::::::::::::::::::::::::::::::::::::::
2746    CTAATGGCTGGATTTTGCCCTCTTCTTGTGGAGTATACCATGGTG 2790

2671    aaaatatatcaaaacagttacacctgggaactttctcctcatttg 2715
        ::::::::::::::::::::::::::::::::::::::::::::
2791    AAAATATATCAAAACAGTTACACCTGGGAACTTTCTCCTCATTTG 2835

2716    tatttccttgctcctggacaacaaccacatgaccctctcactcaa 2760
        ::::::::::::::::::::::::::::::::::::::::::::
2836    TATTTCCTTGCTCCTGGACAACAACCACATGACCCTCTCACTCAA 2880

2761    ctactgatcatcaataaaacaggggcaagcattgatgactttata 2805
        ::::::::::::::::::::::::::::::::::::::::::::
2881    CTACTGATCATCAATAAAACAGGGGCAAGCATTGATGACTTTATA 2925

2806    cagtctgtggagcaccagaacatagctttagaagtggatgcattt 2850
        ::::::::::::::::::::::::::::::::::::::::::::
2926    CAGTCTGTGGAGCACCAGAACATAGCTTTAGAAGTGGATGCATTT 2970

2851    ggaactagaaatggcacagatgacccatcttataatggagccatc 2895
        ::::::::::::::::::::::::::::::::::::::::::::
2971    GGAACTAGAAATGGCACAGATGACCCATCTTATAATGGAGCCATC 3015

2896    acagtgtgttgtaatgaaaagaattacagcttttcgttagcatgc 2940
        ::::::::::::::::::::::::::::::::::::::::::::
3016    ACAGTGTGTTGTAATGAAAAGAATTACAGCTTTTCGTTAGCATGC 3060

2941    aatgccaaaagattgaattgcttcccagttcttatggacattgtt 2985
        ::::::::::::::::::::::::::::::::::::::::::::
3061    AATGCCAAAAGATTGAATTGCTTCCCAGTTCTTATGGACATTGTT 3105

2986    agtaatgggctacttggaatggttaaaccatcagtacatatccga 3030
        ::::::::::::::::::::::::::::::::::::::::::::
3106    AGTAATGGGCTACTTGGAATGGTTAAACCATCAGTACATATCCGA 3150

3031    actgaaagaagtacatttttggagaatggacaggacaatccaatc 3075
        ::::::::::::::::::::::::::::::::::::::::::::
3151    ACTGAAAGAAGTACATTTTTGGAGAATGGACAGGACAATCCAATC 3195

3076    ggattcctggcatatatcatgttctggctggttttaacatcgagt 3120
        ::::::::::::::::::::::::::::::::::::::::::::
3196    GGATTCCTGGCATATATCATGTTCTGGCTGGTTTTAACATCGAGT 3240
```

FIG 4A

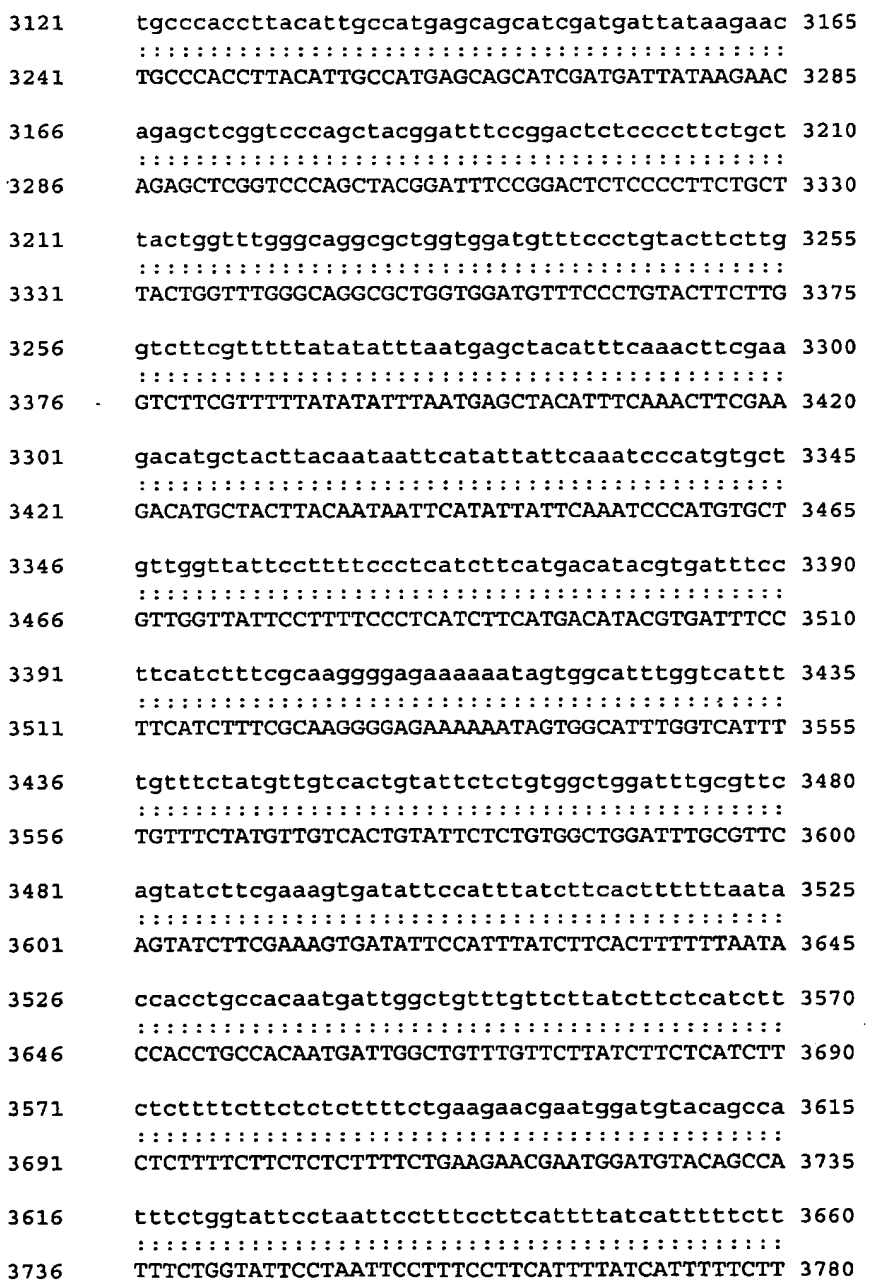
3121 tgcccaccttacattgccatgagcagcatcgatgattataagaac 3165
3241 TGCCCACCTTACATTGCCATGAGCAGCATCGATGATTATAAGAAC 3285
3166 agagctcggtcccagctacggatttccggactctccccttctgct 3210
3286 AGAGCTCGGTCCCAGCTACGGATTTCCGGACTCTCCCCTTCTGCT 3330
3211 tactggtttgggcaggcgctggtggatgtttccctgtacttcttg 3255
3331 TACTGGTTTGGGCAGGCGCTGGTGGATGTTTCCCTGTACTTCTTG 3375
3256 gtcttcgtttttatatatttaatgagctacatttcaaacttcgaa 3300
3376 GTCTTCGTTTTTATATATTTAATGAGCTACATTTCAAACTTCGAA 3420
3301 gacatgctacttacaataattcatattattcaaatcccatgtgct 3345
3421 GACATGCTACTTACAATAATTCATATTATTCAAATCCCATGTGCT 3465
3346 gttggttattccttttccctcatcttcatgacatacgtgatttcc 3390
3466 GTTGGTTATTCCTTTTCCCTCATCTTCATGACATACGTGATTTCC 3510
3391 ttcatctttcgcaaggggagaaaaaatagtggcatttggtcattt 3435
3511 TTCATCTTTCGCAAGGGGAGAAAAAATAGTGGCATTTGGTCATTT 3555
3436 tgtttctatgttgtcactgtattctctgtggctggatttgcgttc 3480
3556 TGTTTCTATGTTGTCACTGTATTCTCTGTGGCTGGATTTGCGTTC 3600
3481 agtatcttcgaaagtgatattccatttatcttcacttttttaata 3525
3601 AGTATCTTCGAAAGTGATATTCCATTTATCTTCACTTTTTTAATA 3645
3526 ccacctgccacaatgattggctgtttgttcttatcttctcatctt 3570
3646 CCACCTGCCACAATGATTGGCTGTTTGTTCTTATCTTCTCATCTT 3690
3571 ctcttttcttctctcttttctgaagaacgaatggatgtacagcca 3615
3691 CTCTTTTCTTCTCTCTTTTCTGAAGAACGAATGGATGTACAGCCA 3735
3616 tttctggtattcctaattcctttccttcattttatcatttttctt 3660
3736 TTTCTGGTATTCCTAATTCCTTTCCTTCATTTTATCATTTTTCTT 3780

FIG 4A

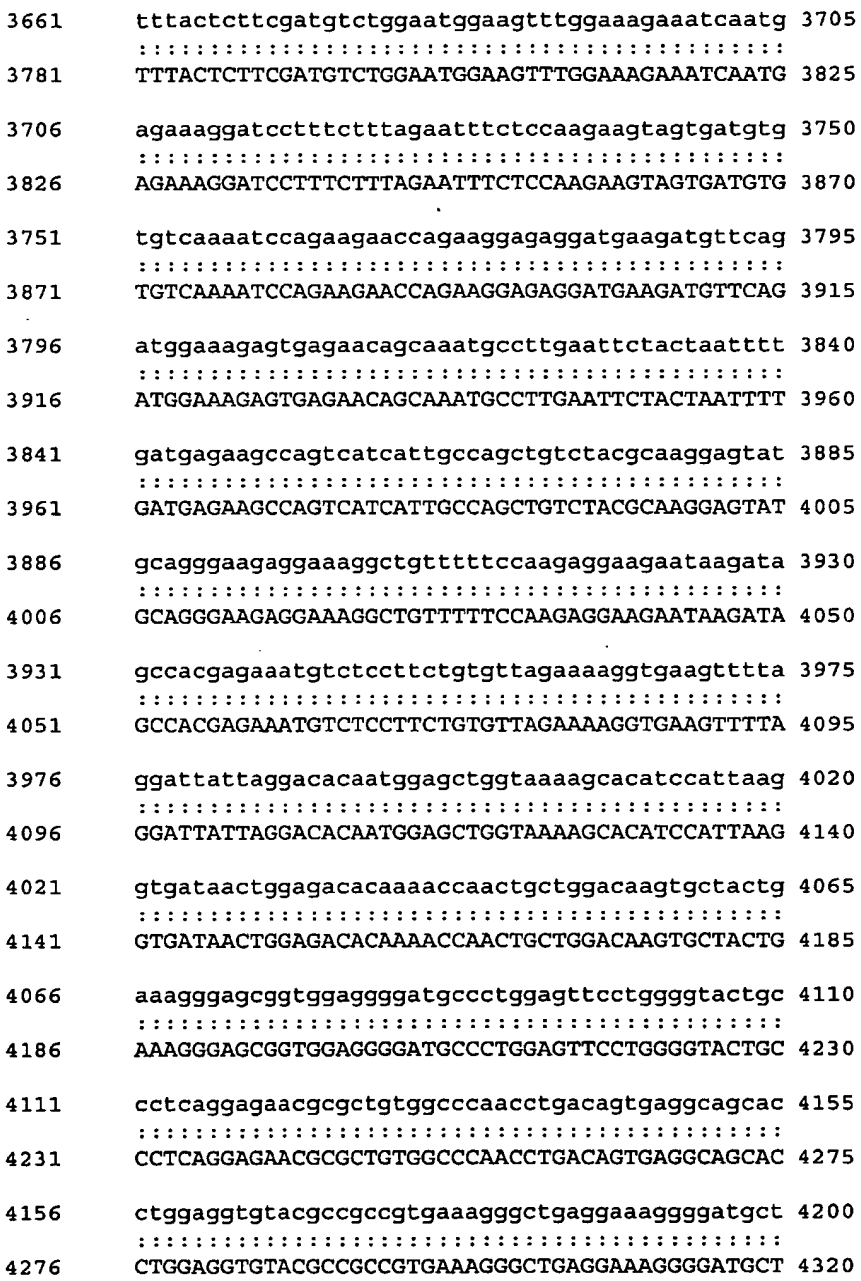
3661 tttactcttcgatgtctggaatggaagtttggaaagaaatcaatg 3705
3781 TTTACTCTTCGATGTCTGGAATGGAAGTTTGGAAAGAAATCAATG 3825
3706 agaaaggatcctttctttagaatttctccaagaagtagtgatgtg 3750
3826 AGAAAGGATCCTTTCTTTAGAATTTCTCCAAGAAGTAGTGATGTG 3870
3751 tgtcaaaatccagaagaaccagaaggagaggatgaagatgttcag 3795
3871 TGTCAAAATCCAGAAGAACCAGAAGGAGAGGATGAAGATGTTCAG 3915
3796 atggaaagagtgagaacagcaaatgccttgaattctactaatttt 3840
3916 ATGGAAAGAGTGAGAACAGCAAATGCCTTGAATTCTACTAATTTT 3960
3841 gatgagaagccagtcatcattgccagctgtctacgcaaggagtat 3885
3961 GATGAGAAGCCAGTCATCATTGCCAGCTGTCTACGCAAGGAGTAT 4005
3886 gcagggaagaggaaaggctgtttttccaagaggaagaataagata 3930
4006 GCAGGGAAGAGGAAAGGCTGTTTTTCCAAGAGGAAGAATAAGATA 4050
3931 gccacgagaaatgtctccttctgtgttagaaaaggtgaagtttta 3975
4051 GCCACGAGAAATGTCTCCTTCTGTGTTAGAAAAGGTGAAGTTTTA 4095
3976 ggattattaggacacaatggagctggtaaaagcacatccattaag 4020
4096 GGATTATTAGGACACAATGGAGCTGGTAAAAGCACATCCATTAAG 4140
4021 gtgataactggagacacaaaaccaactgctggacaagtgctactg 4065
4141 GTGATAACTGGAGACACAAAACCAACTGCTGGACAAGTGCTACTG 4185
4066 aaagggagcggtggaggggatgccctggagttcctggggtactgc 4110
4186 AAAGGGAGCGGTGGAGGGGATGCCCTGGAGTTCCTGGGGTACTGC 4230
4111 cctcaggagaacgcgctgtggcccaacctgacagtgaggcagcac 4155
4231 CCTCAGGAGAACGCGCTGTGGCCCAACCTGACAGTGAGGCAGCAC 4275
4156 ctggaggtgtacgccgccgtgaaagggctgaggaaaggggatgct 4200
4276 CTGGAGGTGTACGCCGCCGTGAAAGGGCTGAGGAAAGGGGATGCT 4320

FIG 4A

```
4201    gaggttgccatcacacggttagtggatgcgctcaagctgcaggac 4245
        :::::::::::::::::::::::::::::::::::::::::::::
4321    GAGGTTGCCATCACACGGTTAGTGGATGCGCTCAAGCTGCAGGAC 4365

4246    cagctgaagtctccggtgaagaccttgtcagagggaataaagaga 4290
        :::::::::::::::::::::::::::::::::::::::::::::
4366    CAGCTGAAGTCTCCGGTGAAGACCTTGTCAGAGGGAATAAAGAGA 4410

4291    aagctgtgctttgtcctgagcatactggggaacccgtcagtggtg 4335
        :::::::::::::::::::::::::::::::::::::::::::::
4411    AAGCTGTGCTTTGTCCTGAGCATACTGGGGAACCCGTCAGTGGTG 4455

4336    cttctggatgagccgtcgaccgggatggaccccgaggggcagcag 4380
        :::::::::::::::::::::::::::::::::::::::::::::
4456    CTTCTGGATGAGCCGTCGACCGGGATGGACCCCGAGGGGCAGCAG 4500

4381    caaatgtggcaggccatccgggccacctttagaaacacggaaagg 4425
        :::::::::::::::::::::::::::::::::::::::::::::
4501    CAAATGTGGCAGGCCATCCGGGCCACCTTTAGAAACACGGAAAGG 4545

4426    ggtgccctcctaaccacccactacatggcagaggctgaggccgtg 4470
        :::::::::::::::::::::::::::::::::::::::::::::
4546    GGTGCCCTCCTAACCACCCACTACATGGCAGAGGCTGAGGCCGTG 4590

4471    tgtgaccgagtggccatcatggtatctgggaggttgagatgtatc 4515
        :::::::::::::::::::::::::::::::::::::::::::::
4591    TGTGACCGAGTGGCCATCATGGTATCTGGGAGGTTGAGATGTATC 4635

4516    ggttccatccaacacctgaaaagcaaatttggcaaagattacctg 4560
        :::::::::::::::::::::::::::::::::::::::::::::
4636    GGTTCCATCCAACACCTGAAAAGCAAATTTGGCAAAGATTACCTG 4680

4561    ctggagatgaaggtgaagaacctggcacaagtggagcccctccat 4605
        :::::::::::::::::::::::::::::::::::::::::::::
4681    CTGGAGATGAAGGTGAAGAACCTGGCACAAGTGGAGCCCCTCCAT 4725

4606    gcagagatcctgaggcttttcccccaggctgctcggcaggaaagg 4650
        :::::::::::::::::::::::::::::::::::::::::::::
4726    GCAGAGATCCTGAGGCTTTTCCCCCAGGCTGCTCGGCAGGAAAGG 4770

4651    tactcctctctgatggtttataagttgccagtggaagatgtgcaa 4695
        :::::::::::::::::::::::::::::::::::::::::::::
4771    TACTCCTCTCTGATGGTTTATAAGTTGCCAGTGGAAGATGTGCAA 4815

4696    cctttagcccaagctttcttcaaattagagaaggttaaacagagc 4740
        :::::::::::::::::::::::::::::::::::::::::::::
4816    CCTTTAGCCCAAGCTTTCTTCAAATTAGAGAAGGTTAAACAGAGC 4860
```

FIG 4A

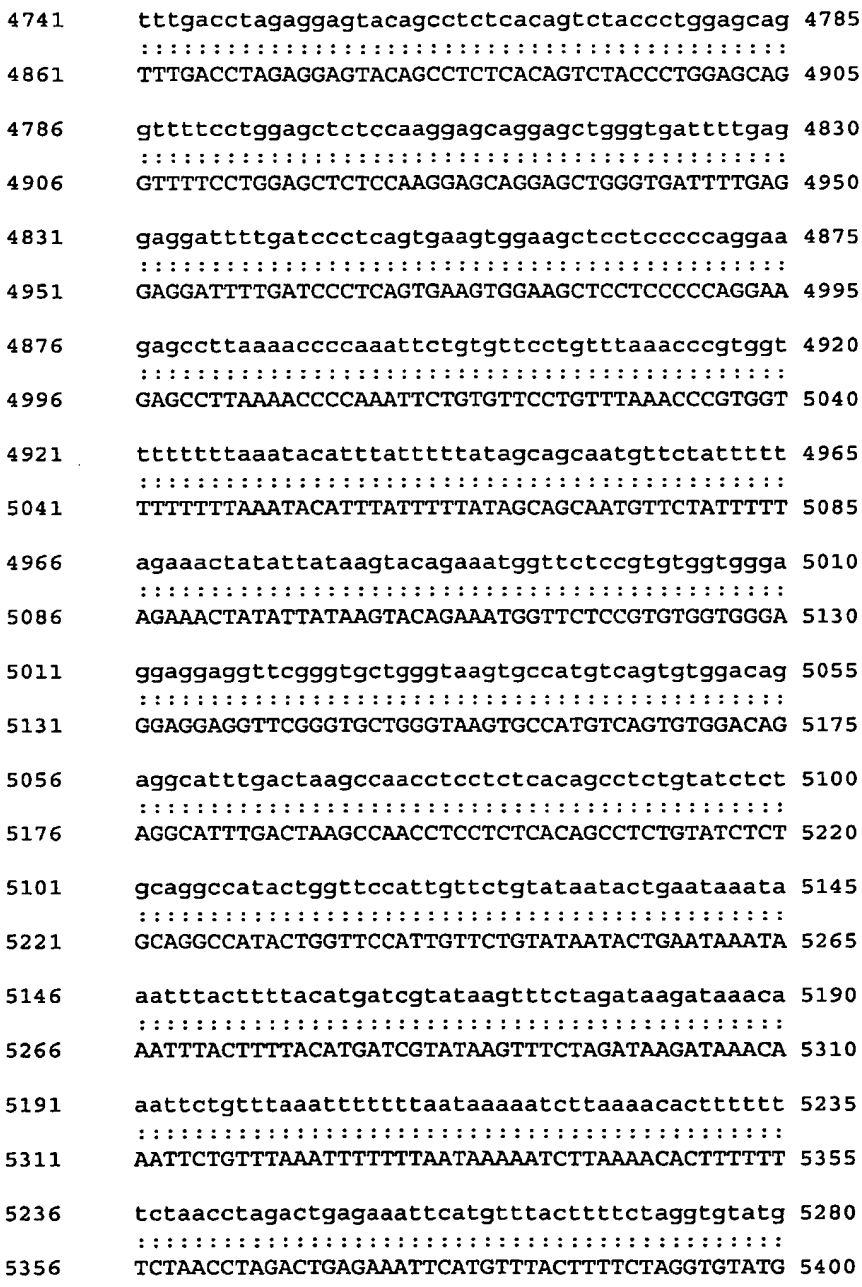
4741 tttgacctagaggagtacagcctctcacagtctaccctggagcag 4785
4861 TTTGACCTAGAGGAGTACAGCCTCTCACAGTCTACCCTGGAGCAG 4905
4786 gttttcctggagctctccaaggagcaggagctgggtgattttgag 4830
4906 GTTTTCCTGGAGCTCTCCAAGGAGCAGGAGCTGGGTGATTTTGAG 4950
4831 gaggattttgatccctcagtgaagtggaagctcctcccccaggaa 4875
4951 GAGGATTTTGATCCCTCAGTGAAGTGGAAGCTCCTCCCCCAGGAA 4995
4876 gagccttaaaaccccaaattctgtgttcctgtttaaacccgtggt 4920
4996 GAGCCTTAAAACCCCAAATTCTGTGTTCCTGTTTAAACCCGTGGT 5040
4921 ttttttaaatacatttatttttatagcagcaatgttctattttt 4965
5041 TTTTTTTAAATACATTTATTTTTATAGCAGCAATGTTCTATTTTT 5085
4966 agaaactatattataagtacagaaatggttctccgtgtggtggga 5010
5086 AGAAACTATATTATAAGTACAGAAATGGTTCTCCGTGTGGTGGGA 5130
5011 ggaggaggttcgggtgctgggtaagtgccatgtcagtgtggacag 5055
5131 GGAGGAGGTTCGGGTGCTGGGTAAGTGCCATGTCAGTGTGGACAG 5175
5056 aggcatttgactaagccaacctcctctcacagcctctgtatctct 5100
5176 AGGCATTTGACTAAGCCAACCTCCTCTCACAGCCTCTGTATCTCT 5220
5101 gcaggccatactggttccattgttctgtataatactgaataaata 5145
5221 GCAGGCCATACTGGTTCCATTGTTCTGTATAATACTGAATAAATA 5265
5146 aatttacttttacatgatcgtataagtttctagataagataaaca 5190
5266 AATTTACTTTTACATGATCGTATAAGTTTCTAGATAAGATAAACA 5310
5191 aattctgtttaaatttttttaataaaaatcttaaaacacttttt 5235
5311 AATTCTGTTTAAATTTTTTTAATAAAAATCTTAAAACACTTTTTT 5355
5236 tctaacctagactgagaaattcatgtttacttttctaggtgtatg 5280
5356 TCTAACCTAGACTGAGAAATTCATGTTTACTTTTCTAGGTGTATG 5400

FIG 4A

```
5281    atactttgtaaagttgatactttcctaagaatttaacatgtcata 5325
        :::::::::::::::::::::::::::::::::::::::::::::
5401    ATACTTTGTAAAGTTGATACTTTCCTAAGAATTTAACATGTCATA 5445

5326    tttttgaaatagatttaagtgtgcttcttattgctaaaaatacta 5370
        :::::::::::::::::::::::::::::::::::::::::::::
5446    TTTTTGAAATAGATTTAAGTGTGCTTCTTATTGCTAAAAATACTA 5490

5371    aatgtcatgggtcatagtatctgatatcaatatcgttgataacat 5415
        :::::::::::::::::::::::::::::::::::::::::::::
5491    AATGTCATGGGTCATAGTATCTGATATCAATATCGTTGATAACAT 5535

5416    atccacaggtaacaccatgatgtaggcataaatggaaaacaaaaa 5460
        :::::::::::::::::::::::::::::::::::::::::::::
5536    ATCCACAGGTAACACCATGATGTAGGCATAAATGGAAAACAAAAA 5580

5461    ccctactatttcaaatatattgtacttttttatttctgtaagcca 5505
        :::::::::::::::::::::::::::::::::::::::::::::
5581    CCCTACTATTTCAAATATATTGTACTTTTTTATTTCTGTAAGCCA 5625

5506    actgtgtgccattttcactggacttttaaatctagactttagtga 5550
        :::::::::::::::::::::::::::::::::::::::::::::
5626    ACTGTGTGCCATTTTCACTGGACTTTTAAATCTAGACTTTAGTGA 5670

5551    tgtctacattgtaaatgatcttttgtggatatttgtcacttggtt 5595
        :::::::::::::::::::::::::::::::::::::::::::::
5671    TGTCTACATTGTAAATGATCTTTTGTGGATATTTGTCACTTGGTT 5715

5596    tcagaaagttcacaaatgtagcaacagctcacatgactgagtagg 5640
        :::::::::::::::::::::::::::::::::::::::::::::
5716    TCAGAAAGTTCACAAATGTAGCAACAGCTCACATGACTGAGTAGG 5760

5641    tagaaaatgtgaaataaatctcatatatatagttttg          5677
        :::::::::::::::::::::::::::::::::::
5761    TAGAAAATGTGAAATAAATCTCATATATATAGTTTTG          579
```

SEQUENCE LISTING

<110> Dean, Michael
Arnould, Isabelle
Prades, Catherine
Allikmets, Rando
Schriml, Lynn
Denefle, Patrice
Rosier, Marie

<120> ABCA8 NUCLEIC ACIDS AND PROTEINS, AND USES THEREOF

<130> 4239-63917

<150> US 60/351,004
<151> 2001-10-19

<160> 10

<170> PatentIn version 3.1

<210> 1
<211> 5797
<212> DNA
<213> homo sapiens

<220>
<221> CDS
<222> (139)..(5004)
<223>

<400> 1
cagaagtcaa atagttaaag caaattctag atacatggta gagaccagga gaaaatatga 60

ataactttct tctaaacaag gagctcagtg gataaaccat acctctagat tccttgcttc 120

cattttccca gaaacaag atg agg aag aga aag atc agt gtg tgt caa caa 171
Met Arg Lys Arg Lys Ile Ser Val Cys Gln Gln
1 5 10

act tgg gcc tta tta tgc aag aac ttt ctt aaa aaa tgg aga atg aaa 219
Thr Trp Ala Leu Leu Cys Lys Asn Phe Leu Lys Lys Trp Arg Met Lys
15 20 25

aga gag tcc tta atg gaa tgg ctg aat tca ttg ctc cta cta ctt tgt 267
Arg Glu Ser Leu Met Glu Trp Leu Asn Ser Leu Leu Leu Leu Leu Cys
30 35 40

ttg tat ata tat cct cat agt cat caa gta aat gat ttt tct tca ctg 315
Leu Tyr Ile Tyr Pro His Ser His Gln Val Asn Asp Phe Ser Ser Leu
45 50 55

ctt acc atg gac ctg gga cgg gta gat aca ttt aat gaa tcc aga ttt 363
Leu Thr Met Asp Leu Gly Arg Val Asp Thr Phe Asn Glu Ser Arg Phe
60 65 70 75

tct gtt gta tac aca cct gtc acc aac acg acc caa cag ata atg aat 411
Ser Val Val Tyr Thr Pro Val Thr Asn Thr Thr Gln Gln Ile Met Asn
80 85 90

```
aaa gta gcc tct act ccc ttc ctg gca ggt aaa gag gtc ttg gga ctg      459
Lys Val Ala Ser Thr Pro Phe Leu Ala Gly Lys Glu Val Leu Gly Leu
            95                  100                 105

cca gat gag gaa agt att aaa gaa ttc aca gca aat tat cct gaa gaa      507
Pro Asp Glu Glu Ser Ile Lys Glu Phe Thr Ala Asn Tyr Pro Glu Glu
        110                 115                 120

ata gta aga gtc acc ttt act aat aca tac tca tat cat ttg aag ttc      555
Ile Val Arg Val Thr Phe Thr Asn Thr Tyr Ser Tyr His Leu Lys Phe
    125                 130                 135

ttg cta gga cat gga atg cca gca aag aag gag cac aag gac cat aca      603
Leu Leu Gly His Gly Met Pro Ala Lys Lys Glu His Lys Asp His Thr
140                 145                 150                 155

gct cat tgt tat gaa aca aat gaa gat gtt tac tgt gaa gtt tca gta      651
Ala His Cys Tyr Glu Thr Asn Glu Asp Val Tyr Cys Glu Val Ser Val
                160                 165                 170

ttt tgg aag gaa ggt ttt gtg gct ctt caa gct gcc att aat gct gct      699
Phe Trp Lys Glu Gly Phe Val Ala Leu Gln Ala Ala Ile Asn Ala Ala
            175                 180                 185

att ata gaa atc aca aca aat cac tca gtg atg gag gag ctg atg tca      747
Ile Ile Glu Ile Thr Thr Asn His Ser Val Met Glu Glu Leu Met Ser
        190                 195                 200

gtt act gga aaa aat atg aag atg cat tcc ttc att ggt caa tca gga      795
Val Thr Gly Lys Asn Met Lys Met His Ser Phe Ile Gly Gln Ser Gly
    205                 210                 215

gtt ata act gat ttg tac ctt ttt tcc tgc att att tca ttt tcc tca      843
Val Ile Thr Asp Leu Tyr Leu Phe Ser Cys Ile Ile Ser Phe Ser Ser
220                 225                 230                 235

ttc att tac tat gca tct gtt aat gtc aca aga gag agg aaa agg atg      891
Phe Ile Tyr Tyr Ala Ser Val Asn Val Thr Arg Glu Arg Lys Arg Met
                240                 245                 250

aag gcc ttg atg aca atg atg ggt ctt cgg gat tca gcg ttc tgg ctc      939
Lys Ala Leu Met Thr Met Met Gly Leu Arg Asp Ser Ala Phe Trp Leu
            255                 260                 265

tcc tgg ggt ttg ctc tat gct ggt ttc atc ttc att atg gcc ctt ttc      987
Ser Trp Gly Leu Leu Tyr Ala Gly Phe Ile Phe Ile Met Ala Leu Phe
        270                 275                 280

ttg gca ctt gtt ata aga tct acc cag ttt atc att ttg tct ggc ttc     1035
Leu Ala Leu Val Ile Arg Ser Thr Gln Phe Ile Ile Leu Ser Gly Phe
    285                 290                 295

atg gta gtc ttc agc ctc ttt ctc ctg tat gga tta tct ttg gta gct     1083
Met Val Val Phe Ser Leu Phe Leu Leu Tyr Gly Leu Ser Leu Val Ala
300                 305                 310                 315

ttg gct ttc tta atg agc atc ttg gta aag aaa tct ttc ctc acc ggc     1131
Leu Ala Phe Leu Met Ser Ile Leu Val Lys Lys Ser Phe Leu Thr Gly
                320                 325                 330

ctg gtc gtg ttc ctc ctc act gtc ttt tgg ggg tgt ctg ggg ttc aca     1179
```

```
Leu Val Val Phe Leu Leu Thr Val Phe Trp Gly Cys Leu Gly Phe Thr
            335                 340                 345

tca ctg tac aga cac ctt cct gca tcc ttg gag tgg att tta agc ttg     1227
Ser Leu Tyr Arg His Leu Pro Ala Ser Leu Glu Trp Ile Leu Ser Leu
        350                 355                 360

ctt agt ccc ttt gcc ttc atg ctt gga atg gcc cag ctt tta cac ttg     1275
Leu Ser Pro Phe Ala Phe Met Leu Gly Met Ala Gln Leu Leu His Leu
    365                 370                 375

gac tat gat ttg aat tct aat gca ttt cct cat cca tcg gac ggc tca     1323
Asp Tyr Asp Leu Asn Ser Asn Ala Phe Pro His Pro Ser Asp Gly Ser
380                 385                 390                 395

aat ctc att gta gca aca aat ttc atg ttg gca ttt gac act tgc ctc     1371
Asn Leu Ile Val Ala Thr Asn Phe Met Leu Ala Phe Asp Thr Cys Leu
                400                 405                 410

tat ctg gca ttg gcg att tac ttt gaa aaa att ttg cca aat gaa tat     1419
Tyr Leu Ala Leu Ala Ile Tyr Phe Glu Lys Ile Leu Pro Asn Glu Tyr
            415                 420                 425

gga cat cga cgt cca cct ttg ttt ttc ctg aag tcc tca ttt tgg tct     1467
Gly His Arg Arg Pro Pro Leu Phe Phe Leu Lys Ser Ser Phe Trp Ser
        430                 435                 440

caa aca caa aag act gat cac gtg gcc ctt gaa gat gaa atg gat gcc     1515
Gln Thr Gln Lys Thr Asp His Val Ala Leu Glu Asp Glu Met Asp Ala
    445                 450                 455

gat cct tca ttt cat gac tct ttt gaa caa gcg cct cca gaa ttc caa     1563
Asp Pro Ser Phe His Asp Ser Phe Glu Gln Ala Pro Pro Glu Phe Gln
460                 465                 470                 475

ggg aaa gaa gcc atc aga atc aga aat gtt aca aaa gaa tat aaa gga     1611
Gly Lys Glu Ala Ile Arg Ile Arg Asn Val Thr Lys Glu Tyr Lys Gly
                480                 485                 490

aag cct gat aaa ata gaa gcc ttg aaa gat ctg gta ttt gac att tac     1659
Lys Pro Asp Lys Ile Glu Ala Leu Lys Asp Leu Val Phe Asp Ile Tyr
            495                 500                 505

gaa ggc caa atc act gca ata ctt ggt cac agt gga gct gga aag tca     1707
Glu Gly Gln Ile Thr Ala Ile Leu Gly His Ser Gly Ala Gly Lys Ser
        510                 515                 520

aca ctg cta aac att ctt agt ggg ttg tct gtt ccc acc aaa ggt tca     1755
Thr Leu Leu Asn Ile Leu Ser Gly Leu Ser Val Pro Thr Lys Gly Ser
    525                 530                 535

gtc acc atc tat aac aat aag ctt tca gaa atg gct gac cta gaa aat     1803
Val Thr Ile Tyr Asn Asn Lys Leu Ser Glu Met Ala Asp Leu Glu Asn
540                 545                 550                 555

ctc agc aag ctg acc gga gtt tgt cca caa tcc aat gtg caa ttt gac     1851
Leu Ser Lys Leu Thr Gly Val Cys Pro Gln Ser Asn Val Gln Phe Asp
                560                 565                 570

ttc ctc act gta aga gaa aac ctc aga ctc ttt gct aaa ata aaa ggg     1899
Phe Leu Thr Val Arg Glu Asn Leu Arg Leu Phe Ala Lys Ile Lys Gly
```

```
            575                 580                 585

att ctg cca caa gaa gtg gat aaa gag ata caa agg gtt ctg ctg gaa     1947
Ile Leu Pro Gln Glu Val Asp Lys Glu Ile Gln Arg Val Leu Leu Glu
        590                 595                 600

ttg gaa atg aaa aat att cag gat gtt ctt gct caa aac tta agt ggt     1995
Leu Glu Met Lys Asn Ile Gln Asp Val Leu Ala Gln Asn Leu Ser Gly
    605                 610                 615

gga cag aaa aga aag cta acc ttt ggg att gcc att tta gga gat cct     2043
Gly Gln Lys Arg Lys Leu Thr Phe Gly Ile Ala Ile Leu Gly Asp Pro
620                 625                 630                 635

cag att ttc ctg ttg gat gaa cca act gct gga ttg gat ccc ttt tca     2091
Gln Ile Phe Leu Leu Asp Glu Pro Thr Ala Gly Leu Asp Pro Phe Ser
                640                 645                 650

aga cac caa gta tgg aac ctt ctg aaa gaa cgc aaa aca gac cgc gtg     2139
Arg His Gln Val Trp Asn Leu Leu Lys Glu Arg Lys Thr Asp Arg Val
            655                 660                 665

atc ctc ttc agt acc cag ttc atg gat gag gcc gac atc ctg gcg gac     2187
Ile Leu Phe Ser Thr Gln Phe Met Asp Glu Ala Asp Ile Leu Ala Asp
        670                 675                 680

agg aaa gta ttt ctc tcc caa ggg aag cta aag tgc gcg ggc tct tct     2235
Arg Lys Val Phe Leu Ser Gln Gly Lys Leu Lys Cys Ala Gly Ser Ser
    685                 690                 695

ttg ttt cta aag aag aaa tgg ggg att gga tat cac tta agc ttg cag     2283
Leu Phe Leu Lys Lys Lys Trp Gly Ile Gly Tyr His Leu Ser Leu Gln
700                 705                 710                 715

tta aat gaa ata tgt gtt gag gaa aac ata aca tca ctt gtt aaa cag     2331
Leu Asn Glu Ile Cys Val Glu Glu Asn Ile Thr Ser Leu Val Lys Gln
                720                 725                 730

cac atc cct gat gcc aaa tta tca gcc aaa agc gaa gga aaa ctt att     2379
His Ile Pro Asp Ala Lys Leu Ser Ala Lys Ser Glu Gly Lys Leu Ile
            735                 740                 745

tat aca tta ccc tta gaa aga aca aat aaa ttt cca gaa ctt tac aag     2427
Tyr Thr Leu Pro Leu Glu Arg Thr Asn Lys Phe Pro Glu Leu Tyr Lys
        750                 755                 760

gat ctt gat agc tat cct gac cta gga att gag aat tat ggt gtt tcc     2475
Asp Leu Asp Ser Tyr Pro Asp Leu Gly Ile Glu Asn Tyr Gly Val Ser
    765                 770                 775

atg aca act ttg aat gaa gta ttc ctg aag cta gaa gga aaa tct aca     2523
Met Thr Thr Leu Asn Glu Val Phe Leu Lys Leu Glu Gly Lys Ser Thr
780                 785                 790                 795

att aat gaa tcg gac att gct att ttg gga gaa gta caa gcg gaa aaa     2571
Ile Asn Glu Ser Asp Ile Ala Ile Leu Gly Glu Val Gln Ala Glu Lys
                800                 805                 810

gct gac gac act gaa agg ctt gtt gag atg gaa caa gtc ctc tct tca     2619
Ala Asp Asp Thr Glu Arg Leu Val Glu Met Glu Gln Val Leu Ser Ser
            815                 820                 825
```

```
ctt aac aag atg aga aag aca ata ggt ggt gtg gct ctc tgg cga cag     2667
Leu Asn Lys Met Arg Lys Thr Ile Gly Gly Val Ala Leu Trp Arg Gln
        830                 835                 840

caa atc tgc gca att gca agg gtt cgc ttg tta aag tta aag cat gaa     2715
Gln Ile Cys Ala Ile Ala Arg Val Arg Leu Leu Lys Leu Lys His Glu
    845                 850                 855

aga aaa gct ctt tta gca ctg cta tta att cta atg gct gga ttt tgc     2763
Arg Lys Ala Leu Leu Ala Leu Leu Leu Ile Leu Met Ala Gly Phe Cys
860                 865                 870                 875

cct ctt ctt gtg gag tat acc atg gtg aaa ata tat caa aac agt tac     2811
Pro Leu Leu Val Glu Tyr Thr Met Val Lys Ile Tyr Gln Asn Ser Tyr
                880                 885                 890

acc tgg gaa ctt tct cct cat ttg tat ttc ctt gct cct gga caa caa     2859
Thr Trp Glu Leu Ser Pro His Leu Tyr Phe Leu Ala Pro Gly Gln Gln
            895                 900                 905

cca cat gac cct ctc act caa cta ctg atc atc aat aaa aca ggg gca     2907
Pro His Asp Pro Leu Thr Gln Leu Leu Ile Ile Asn Lys Thr Gly Ala
        910                 915                 920

agc att gat gac ttt ata cag tct gtg gag cac cag aac ata gct tta     2955
Ser Ile Asp Asp Phe Ile Gln Ser Val Glu His Gln Asn Ile Ala Leu
    925                 930                 935

gaa gtg gat gca ttt gga act aga aat ggc aca gat gac cca tct tat     3003
Glu Val Asp Ala Phe Gly Thr Arg Asn Gly Thr Asp Asp Pro Ser Tyr
940                 945                 950                 955

aat gga gcc atc aca gtg tgt tgt aat gaa aag aat tac agc ttt tcg     3051
Asn Gly Ala Ile Thr Val Cys Cys Asn Glu Lys Asn Tyr Ser Phe Ser
                960                 965                 970

tta gca tgc aat gcc aaa aga ttg aat tgc ttc cca gtt ctt atg gac     3099
Leu Ala Cys Asn Ala Lys Arg Leu Asn Cys Phe Pro Val Leu Met Asp
            975                 980                 985

att gtt agt aat ggg cta ctt gga atg gtt aaa cca tca  gta cat atc    3147
Ile Val Ser Asn Gly Leu Leu Gly Met Val Lys Pro Ser  Val His Ile
        990                 995                 1000

cga act  gaa aga agt aca ttt  ttg gag aat gga cag  gac aat cca      3192
Arg Thr  Glu Arg Ser Thr Phe  Leu Glu Asn Gly Gln  Asp Asn Pro
    1005                1010                1015

atc gga  ttc ctg gca tat atc  atg ttc tgg ctg gtt  tta aca tcg      3237
Ile Gly  Phe Leu Ala Tyr Ile  Met Phe Trp Leu Val  Leu Thr Ser
    1020                1025                1030

agt tgc  cca cct tac att gcc  atg agc agc atc gat  gat tat aag      3282
Ser Cys  Pro Pro Tyr Ile Ala  Met Ser Ser Ile Asp  Asp Tyr Lys
    1035                1040                1045

aac aga  gct cgg tcc cag cta  cgg att tcc gga ctc  tcc cct tct      3327
Asn Arg  Ala Arg Ser Gln Leu  Arg Ile Ser Gly Leu  Ser Pro Ser
    1050                1055                1060
```

```
gct tac  tgg ttt ggg cag gcg  ctg gtg gat gtt tcc  ctg tac ttc      3372
Ala Tyr  Trp Phe Gly Gln Ala  Leu Val Asp Val Ser  Leu Tyr Phe
    1065                 1070                 1075

ttg gtc  ttc gtt ttt ata tat  tta atg agc tac att  tca aac ttc      3417
Leu Val  Phe Val Phe Ile Tyr  Leu Met Ser Tyr Ile  Ser Asn Phe
    1080                 1085                 1090

gaa gac  atg cta ctt aca ata  att cat att att caa  atc cca tgt      3462
Glu Asp  Met Leu Leu Thr Ile  Ile His Ile Ile Gln  Ile Pro Cys
    1095                 1100                 1105

gct gtt  ggt tat tcc ttt tcc  ctc atc ttc atg aca  tac gtg att      3507
Ala Val  Gly Tyr Ser Phe Ser  Leu Ile Phe Met Thr  Tyr Val Ile
    1110                 1115                 1120

tcc ttc  atc ttt cgc aag ggg  aga aaa aat agt ggc  att tgg tca      3552
Ser Phe  Ile Phe Arg Lys Gly  Arg Lys Asn Ser Gly  Ile Trp Ser
    1125                 1130                 1135

ttt tgt  ttc tat gtt gtc act  gta ttc tct gtg gct  gga ttt gcg      3597
Phe Cys  Phe Tyr Val Val Thr  Val Phe Ser Val Ala  Gly Phe Ala
    1140                 1145                 1150

ttc agt  atc ttc gaa agt gat  att cca ttt atc ttc  act ttt tta      3642
Phe Ser  Ile Phe Glu Ser Asp  Ile Pro Phe Ile Phe  Thr Phe Leu
    1155                 1160                 1165

ata cca  cct gcc aca atg att  ggc tgt ttg ttc tta  tct tct cat      3687
Ile Pro  Pro Ala Thr Met Ile  Gly Cys Leu Phe Leu  Ser Ser His
    1170                 1175                 1180

ctt ctc  ttt tct tct ctc ttt  tct gaa gaa cga atg  gat gta cag      3732
Leu Leu  Phe Ser Ser Leu Phe  Ser Glu Glu Arg Met  Asp Val Gln
    1185                 1190                 1195

cca ttt  ctg gta ttc cta att  cct ttc ctt cat ttt  atc att ttt      3777
Pro Phe  Leu Val Phe Leu Ile  Pro Phe Leu His Phe  Ile Ile Phe
    1200                 1205                 1210

ctt ttt  act ctt cga tgt ctg  gaa tgg aag ttt gga  aag aaa tca      3822
Leu Phe  Thr Leu Arg Cys Leu  Glu Trp Lys Phe Gly  Lys Lys Ser
    1215                 1220                 1225

atg aga  aag gat cct ttc ttt  aga att tct cca aga  agt agt gat      3867
Met Arg  Lys Asp Pro Phe Phe  Arg Ile Ser Pro Arg  Ser Ser Asp
    1230                 1235                 1240

gtg tgt  caa aat cca gaa gaa  cca gaa gga gag gat  gaa gat gtt      3912
Val Cys  Gln Asn Pro Glu Glu  Pro Glu Gly Glu Asp  Glu Asp Val
    1245                 1250                 1255

cag atg  gaa aga gtg aga aca  gca aat gcc ttg aat  tct act aat      3957
Gln Met  Glu Arg Val Arg Thr  Ala Asn Ala Leu Asn  Ser Thr Asn
    1260                 1265                 1270

ttt gat  gag aag cca gtc atc  att gcc agc tgt cta  cgc aag gag      4002
Phe Asp  Glu Lys Pro Val Ile  Ile Ala Ser Cys Leu  Arg Lys Glu
    1275                 1280                 1285

tat gca  ggg aag agg aaa ggc  tgt ttt tcc aag agg  aag aat aag      4047
```

```
Tyr Ala  Gly Lys Arg Lys Gly  Cys Phe Ser Lys Arg  Lys Asn Lys
    1290                 1295                 1300

ata gcc  acg aga aat gtc tcc  ttc tgt gtt aga aaa  ggt gaa gtt      4092
Ile Ala  Thr Arg Asn Val Ser  Phe Cys Val Arg Lys  Gly Glu Val
    1305                 1310                 1315

tta gga  tta tta gga cac aat  gga gct ggt aaa agc  aca tcc att      4137
Leu Gly  Leu Leu Gly His Asn  Gly Ala Gly Lys Ser  Thr Ser Ile
    1320                 1325                 1330

aag gtg  ata act gga gac aca  aaa cca act gct gga  caa gtg cta      4182
Lys Val  Ile Thr Gly Asp Thr  Lys Pro Thr Ala Gly  Gln Val Leu
    1335                 1340                 1345

ctg aaa  ggg agc ggt gga ggg  gat gcc ctg gag ttc  ctg ggg tac      4227
Leu Lys  Gly Ser Gly Gly Gly  Asp Ala Leu Glu Phe  Leu Gly Tyr
    1350                 1355                 1360

tgc cct  cag gag aac gcg ctg  tgg ccc aac ctg aca  gtg agg cag      4272
Cys Pro  Gln Glu Asn Ala Leu  Trp Pro Asn Leu Thr  Val Arg Gln
    1365                 1370                 1375

cac ctg  gag gtg tac gcc gcc  gtg aaa ggg ctg agg  aaa ggg gat      4317
His Leu  Glu Val Tyr Ala Ala  Val Lys Gly Leu Arg  Lys Gly Asp
    1380                 1385                 1390

gct gag  gtt gcc atc aca cgg  tta gtg gat gcg ctc  aag ctg cag      4362
Ala Glu  Val Ala Ile Thr Arg  Leu Val Asp Ala Leu  Lys Leu Gln
    1395                 1400                 1405

gac cag  ctg aag tct ccg gtg  aag acc ttg tca gag  gga ata aag      4407
Asp Gln  Leu Lys Ser Pro Val  Lys Thr Leu Ser Glu  Gly Ile Lys
    1410                 1415                 1420

aga aag  ctg tgc ttt gtc ctg  agc ata ctg ggg aac  ccg tca gtg      4452
Arg Lys  Leu Cys Phe Val Leu  Ser Ile Leu Gly Asn  Pro Ser Val
    1425                 1430                 1435

gtg ctt  ctg gat gag ccg tcg  acc ggg atg gac ccc  gag ggg cag      4497
Val Leu  Leu Asp Glu Pro Ser  Thr Gly Met Asp Pro  Glu Gly Gln
    1440                 1445                 1450

cag caa  atg tgg cag gcc atc  cgg gcc acc ttt aga  aac acg gaa      4542
Gln Gln  Met Trp Gln Ala Ile  Arg Ala Thr Phe Arg  Asn Thr Glu
    1455                 1460                 1465

agg ggt  gcc ctc cta acc acc  cac tac atg gca gag  gct gag gcc      4587
Arg Gly  Ala Leu Leu Thr Thr  His Tyr Met Ala Glu  Ala Glu Ala
    1470                 1475                 1480

gtg tgt  gac cga gtg gcc atc  atg gta tct ggg agg  ttg aga tgt      4632
Val Cys  Asp Arg Val Ala Ile  Met Val Ser Gly Arg  Leu Arg Cys
    1485                 1490                 1495

atc ggt  tcc atc caa cac ctg  aaa agc aaa ttt ggc  aaa gat tac      4677
Ile Gly  Ser Ile Gln His Leu  Lys Ser Lys Phe Gly  Lys Asp Tyr
    1500                 1505                 1510

ctg ctg  gag atg aag gtg aag  aac ctg gca caa gtg  gag ccc ctc      4722
Leu Leu  Glu Met Lys Val Lys  Asn Leu Ala Gln Val  Glu Pro Leu
```

```
    1515                1520                1525

cat gca  gag atc ctg agg ctt  ttc ccc cag gct gct  cgg cag gaa      4767
His Ala  Glu Ile Leu Arg Leu  Phe Pro Gln Ala Ala  Arg Gln Glu
    1530                1535                1540

agg tac  tcc tct ctg atg gtt  tat aag ttg cca gtg  gaa gat gtg      4812
Arg Tyr  Ser Ser Leu Met Val  Tyr Lys Leu Pro Val  Glu Asp Val
    1545                1550                1555

caa cct  tta gcc caa gct ttc  ttc aaa tta gag aag  gtt aaa cag      4857
Gln Pro  Leu Ala Gln Ala Phe  Phe Lys Leu Glu Lys  Val Lys Gln
    1560                1565                1570

agc ttt  gac cta gag gag tac  agc ctc tca cag tct  acc ctg gag      4902
Ser Phe  Asp Leu Glu Glu Tyr  Ser Leu Ser Gln Ser  Thr Leu Glu
    1575                1580                1585

cag gtt  ttc ctg gag ctc tcc  aag gag cag gag ctg  ggt gat ttt      4947
Gln Val  Phe Leu Glu Leu Ser  Lys Glu Gln Glu Leu  Gly Asp Phe
    1590                1595                1600

gag gag  gat ttt gat ccc tca  gtg aag tgg aag ctc  ctc ccc cag      4992
Glu Glu  Asp Phe Asp Pro Ser  Val Lys Trp Lys Leu  Leu Pro Gln
    1605                1610                1615

gaa gag  cct taa aaccccaaat tctgtgttcc tgtttaaacc cgtggttttt       5044
Glu Glu  Pro
    1620

tttaaataca tttattttta tagcagcaat gttctatttt tagaaactat attataagta  5104

cagaaatggt tctccgtgtg gtgggaggag gaggttcggg tgctgggtaa gtgccatgtc  5164

agtgtggaca gaggcatttg actaagccaa cctcctctca cagcctctgt atctctgcag  5224

gccatactgg ttccattgtt ctgtataata ctgaataaat aaatttactt ttacatgatc  5284

gtataagttt ctagataaga taaacaaatt ctgtttaaat ttttttaata aaaatcttaa  5344

aacacttttt ttctaaccta gactgagaaa ttcatgttta cttttctagg tgtatgatac  5404

tttgtaaagt tgatactttc ctaagaattt aacatgtcat atttttgaaa tagatttaag  5464

tgtgcttctt attgctaaaa atactaaatg tcatgggtca tagtatctga tatcaatatc  5524

gttgataaca tatccacagg taacaccatg atgtaggcat aaatggaaaa caaaaaccct  5584

actatttcaa atatattgta ctttttttatt tctgtaagcc aactgtgtgc cattttcact  5644

ggacttttaa atctagactt tagtgatgtc tacattgtaa atgatctttt gtggatattt  5704

gtcacttggt ttcagaaagt tcacaaatgt agcaacagct cacatgactg agtaggtaga  5764

aaatgtgaaa taaatctcat atatatagtt ttg                               5797


<210>  2
<211>  1621
<212>  PRT
<213>  homo sapiens
```

<400> 2

```
Met Arg Lys Arg Lys Ile Ser Val Cys Gln Gln Thr Trp Ala Leu Leu
1               5                  10                  15

Cys Lys Asn Phe Leu Lys Lys Trp Arg Met Lys Arg Glu Ser Leu Met
            20                  25                  30

Glu Trp Leu Asn Ser Leu Leu Leu Leu Leu Cys Leu Tyr Ile Tyr Pro
        35                  40                  45

His Ser His Gln Val Asn Asp Phe Ser Ser Leu Leu Thr Met Asp Leu
    50                  55                  60

Gly Arg Val Asp Thr Phe Asn Glu Ser Arg Phe Ser Val Val Tyr Thr
65                  70                  75                  80

Pro Val Thr Asn Thr Thr Gln Gln Ile Met Asn Lys Val Ala Ser Thr
                85                  90                  95

Pro Phe Leu Ala Gly Lys Glu Val Leu Gly Leu Pro Asp Glu Glu Ser
            100                 105                 110

Ile Lys Glu Phe Thr Ala Asn Tyr Pro Glu Glu Ile Val Arg Val Thr
        115                 120                 125

Phe Thr Asn Thr Tyr Ser Tyr His Leu Lys Phe Leu Leu Gly His Gly
    130                 135                 140

Met Pro Ala Lys Lys Glu His Lys Asp His Thr Ala His Cys Tyr Glu
145                 150                 155                 160

Thr Asn Glu Asp Val Tyr Cys Glu Val Ser Val Phe Trp Lys Glu Gly
                165                 170                 175

Phe Val Ala Leu Gln Ala Ala Ile Asn Ala Ala Ile Ile Glu Ile Thr
            180                 185                 190

Thr Asn His Ser Val Met Glu Glu Leu Met Ser Val Thr Gly Lys Asn
        195                 200                 205

Met Lys Met His Ser Phe Ile Gly Gln Ser Gly Val Ile Thr Asp Leu
    210                 215                 220

Tyr Leu Phe Ser Cys Ile Ile Ser Phe Ser Ser Phe Ile Tyr Tyr Ala
225                 230                 235                 240
```

```
Ser Val Asn Val Thr Arg Glu Arg Lys Arg Met Lys Ala Leu Met Thr
                245                 250                 255

Met Met Gly Leu Arg Asp Ser Ala Phe Trp Leu Ser Trp Gly Leu Leu
            260                 265                 270

Tyr Ala Gly Phe Ile Phe Ile Met Ala Leu Phe Leu Ala Leu Val Ile
        275                 280                 285

Arg Ser Thr Gln Phe Ile Ile Leu Ser Gly Phe Met Val Val Phe Ser
    290                 295                 300

Leu Phe Leu Leu Tyr Gly Leu Ser Leu Val Ala Leu Ala Phe Leu Met
305                 310                 315                 320

Ser Ile Leu Val Lys Lys Ser Phe Leu Thr Gly Leu Val Val Phe Leu
                325                 330                 335

Leu Thr Val Phe Trp Gly Cys Leu Gly Phe Thr Ser Leu Tyr Arg His
            340                 345                 350

Leu Pro Ala Ser Leu Glu Trp Ile Leu Ser Leu Leu Ser Pro Phe Ala
        355                 360                 365

Phe Met Leu Gly Met Ala Gln Leu Leu His Leu Asp Tyr Asp Leu Asn
    370                 375                 380

Ser Asn Ala Phe Pro His Pro Ser Asp Gly Ser Asn Leu Ile Val Ala
385                 390                 395                 400

Thr Asn Phe Met Leu Ala Phe Asp Thr Cys Leu Tyr Leu Ala Leu Ala
                405                 410                 415

Ile Tyr Phe Glu Lys Ile Leu Pro Asn Glu Tyr Gly His Arg Arg Pro
            420                 425                 430

Pro Leu Phe Phe Leu Lys Ser Ser Phe Trp Ser Gln Thr Gln Lys Thr
        435                 440                 445

Asp His Val Ala Leu Glu Asp Glu Met Asp Ala Asp Pro Ser Phe His
    450                 455                 460

Asp Ser Phe Glu Gln Ala Pro Pro Glu Phe Gln Gly Lys Glu Ala Ile
465                 470                 475                 480
```

```
Arg Ile Arg Asn Val Thr Lys Glu Tyr Lys Gly Lys Pro Asp Lys Ile
                485                 490                 495

Glu Ala Leu Lys Asp Leu Val Phe Asp Ile Tyr Glu Gly Gln Ile Thr
            500                 505                 510

Ala Ile Leu Gly His Ser Gly Ala Gly Lys Ser Thr Leu Leu Asn Ile
        515                 520                 525

Leu Ser Gly Leu Ser Val Pro Thr Lys Gly Ser Val Thr Ile Tyr Asn
    530                 535                 540

Asn Lys Leu Ser Glu Met Ala Asp Leu Glu Asn Leu Ser Lys Leu Thr
545                 550                 555                 560

Gly Val Cys Pro Gln Ser Asn Val Gln Phe Asp Phe Leu Thr Val Arg
                565                 570                 575

Glu Asn Leu Arg Leu Phe Ala Lys Ile Lys Gly Ile Leu Pro Gln Glu
            580                 585                 590

Val Asp Lys Glu Ile Gln Arg Val Leu Leu Glu Leu Glu Met Lys Asn
        595                 600                 605

Ile Gln Asp Val Leu Ala Gln Asn Leu Ser Gly Gly Gln Lys Arg Lys
    610                 615                 620

Leu Thr Phe Gly Ile Ala Ile Leu Gly Asp Pro Gln Ile Phe Leu Leu
625                 630                 635                 640

Asp Glu Pro Thr Ala Gly Leu Asp Pro Phe Ser Arg His Gln Val Trp
                645                 650                 655

Asn Leu Leu Lys Glu Arg Lys Thr Asp Arg Val Ile Leu Phe Ser Thr
            660                 665                 670

Gln Phe Met Asp Glu Ala Asp Ile Leu Ala Asp Arg Lys Val Phe Leu
        675                 680                 685

Ser Gln Gly Lys Leu Lys Cys Ala Gly Ser Ser Leu Phe Leu Lys Lys
    690                 695                 700

Lys Trp Gly Ile Gly Tyr His Leu Ser Leu Gln Leu Asn Glu Ile Cys
705                 710                 715                 720
```

```
Val Glu Glu Asn Ile Thr Ser Leu Val Lys Gln His Ile Pro Asp Ala
                725                 730                 735

Lys Leu Ser Ala Lys Ser Glu Gly Lys Leu Ile Tyr Thr Leu Pro Leu
            740                 745                 750

Glu Arg Thr Asn Lys Phe Pro Glu Leu Tyr Lys Asp Leu Asp Ser Tyr
        755                 760                 765

Pro Asp Leu Gly Ile Glu Asn Tyr Gly Val Ser Met Thr Thr Leu Asn
    770                 775                 780

Glu Val Phe Leu Lys Leu Glu Gly Lys Ser Thr Ile Asn Glu Ser Asp
785                 790                 795                 800

Ile Ala Ile Leu Gly Glu Val Gln Ala Glu Lys Ala Asp Asp Thr Glu
                805                 810                 815

Arg Leu Val Glu Met Glu Gln Val Leu Ser Ser Leu Asn Lys Met Arg
            820                 825                 830

Lys Thr Ile Gly Gly Val Ala Leu Trp Arg Gln Gln Ile Cys Ala Ile
        835                 840                 845

Ala Arg Val Arg Leu Leu Lys Leu Lys His Glu Arg Lys Ala Leu Leu
    850                 855                 860

Ala Leu Leu Leu Ile Leu Met Ala Gly Phe Cys Pro Leu Leu Val Glu
865                 870                 875                 880

Tyr Thr Met Val Lys Ile Tyr Gln Asn Ser Tyr Thr Trp Glu Leu Ser
                885                 890                 895

Pro His Leu Tyr Phe Leu Ala Pro Gly Gln Gln Pro His Asp Pro Leu
            900                 905                 910

Thr Gln Leu Leu Ile Ile Asn Lys Thr Gly Ala Ser Ile Asp Asp Phe
        915                 920                 925

Ile Gln Ser Val Glu His Gln Asn Ile Ala Leu Glu Val Asp Ala Phe
    930                 935                 940

Gly Thr Arg Asn Gly Thr Asp Asp Pro Ser Tyr Asn Gly Ala Ile Thr
945                 950                 955                 960

Val Cys Cys Asn Glu Lys Asn Tyr Ser Phe Ser Leu Ala Cys Asn Ala
```

965 970 975

Lys Arg Leu Asn Cys Phe Pro Val Leu Met Asp Ile Val Ser Asn Gly
980 985 990

Leu Leu Gly Met Val Lys Pro Ser Val His Ile Arg Thr Glu Arg Ser
995 1000 1005

Thr Phe Leu Glu Asn Gly Gln Asp Asn Pro Ile Gly Phe Leu Ala
1010 1015 1020

Tyr Ile Met Phe Trp Leu Val Leu Thr Ser Ser Cys Pro Pro Tyr
1025 1030 1035

Ile Ala Met Ser Ser Ile Asp Asp Tyr Lys Asn Arg Ala Arg Ser
1040 1045 1050

Gln Leu Arg Ile Ser Gly Leu Ser Pro Ser Ala Tyr Trp Phe Gly
1055 1060 1065

Gln Ala Leu Val Asp Val Ser Leu Tyr Phe Leu Val Phe Val Phe
1070 1075 1080

Ile Tyr Leu Met Ser Tyr Ile Ser Asn Phe Glu Asp Met Leu Leu
1085 1090 1095

Thr Ile Ile His Ile Ile Gln Ile Pro Cys Ala Val Gly Tyr Ser
1100 1105 1110

Phe Ser Leu Ile Phe Met Thr Tyr Val Ile Ser Phe Ile Phe Arg
1115 1120 1125

Lys Gly Arg Lys Asn Ser Gly Ile Trp Ser Phe Cys Phe Tyr Val
1130 1135 1140

Val Thr Val Phe Ser Val Ala Gly Phe Ala Phe Ser Ile Phe Glu
1145 1150 1155

Ser Asp Ile Pro Phe Ile Phe Thr Phe Leu Ile Pro Pro Ala Thr
1160 1165 1170

Met Ile Gly Cys Leu Phe Leu Ser Ser His Leu Leu Phe Ser Ser
1175 1180 1185

Leu Phe Ser Glu Glu Arg Met Asp Val Gln Pro Phe Leu Val Phe
1190 1195 1200

```
Leu Ile  Pro Phe Leu His Phe  Ile Ile Phe Leu Phe  Thr Leu Arg
    1205                 1210                 1215

Cys Leu  Glu Trp Lys Phe Gly  Lys Lys Ser Met Arg  Lys Asp Pro
    1220                 1225                 1230

Phe Phe  Arg Ile Ser Pro Arg  Ser Ser Asp Val Cys  Gln Asn Pro
    1235                 1240                 1245

Glu Glu  Pro Glu Gly Glu Asp  Glu Asp Val Gln Met  Glu Arg Val
    1250                 1255                 1260

Arg Thr  Ala Asn Ala Leu Asn  Ser Thr Asn Phe Asp  Glu Lys Pro
    1265                 1270                 1275

Val Ile  Ile Ala Ser Cys Leu  Arg Lys Glu Tyr Ala  Gly Lys Arg
    1280                 1285                 1290

Lys Gly  Cys Phe Ser Lys Arg  Lys Asn Lys Ile Ala  Thr Arg Asn
    1295                 1300                 1305

Val Ser  Phe Cys Val Arg Lys  Gly Glu Val Leu Gly  Leu Leu Gly
    1310                 1315                 1320

His Asn  Gly Ala Gly Lys Ser  Thr Ser Ile Lys Val  Ile Thr Gly
    1325                 1330                 1335

Asp Thr  Lys Pro Thr Ala Gly  Gln Val Leu Leu Lys  Gly Ser Gly
    1340                 1345                 1350

Gly Gly  Asp Ala Leu Glu Phe  Leu Gly Tyr Cys Pro  Gln Glu Asn
    1355                 1360                 1365

Ala Leu  Trp Pro Asn Leu Thr  Val Arg Gln His Leu  Glu Val Tyr
    1370                 1375                 1380

Ala Ala  Val Lys Gly Leu Arg  Lys Gly Asp Ala Glu  Val Ala Ile
    1385                 1390                 1395

Thr Arg  Leu Val Asp Ala Leu  Lys Leu Gln Asp Gln  Leu Lys Ser
    1400                 1405                 1410

Pro Val  Lys Thr Leu Ser Glu  Gly Ile Lys Arg Lys  Leu Cys Phe
    1415                 1420                 1425
```

```
Val Leu  Ser Ile Leu Gly Asn  Pro Ser Val Val Leu  Leu Asp Glu
    1430                 1435                 1440

Pro Ser  Thr Gly Met Asp Pro  Glu Gly Gln Gln Gln  Met Trp Gln
    1445                 1450                 1455

Ala Ile  Arg Ala Thr Phe Arg  Asn Thr Glu Arg Gly  Ala Leu Leu
    1460                 1465                 1470

Thr Thr  His Tyr Met Ala Glu  Ala Glu Ala Val Cys  Asp Arg Val
    1475                 1480                 1485

Ala Ile  Met Val Ser Gly Arg  Leu Arg Cys Ile Gly  Ser Ile Gln
    1490                 1495                 1500

His Leu  Lys Ser Lys Phe Gly  Lys Asp Tyr Leu Leu  Glu Met Lys
    1505                 1510                 1515

Val Lys  Asn Leu Ala Gln Val  Glu Pro Leu His Ala  Glu Ile Leu
    1520                 1525                 1530

Arg Leu  Phe Pro Gln Ala Ala  Arg Gln Glu Arg Tyr  Ser Ser Leu
    1535                 1540                 1545

Met Val  Tyr Lys Leu Pro Val  Glu Asp Val Gln Pro  Leu Ala Gln
    1550                 1555                 1560

Ala Phe  Phe Lys Leu Glu Lys  Val Lys Gln Ser Phe  Asp Leu Glu
    1565                 1570                 1575

Glu Tyr  Ser Leu Ser Gln Ser  Thr Leu Glu Gln Val  Phe Leu Glu
    1580                 1585                 1590

Leu Ser  Lys Glu Gln Glu Leu  Gly Asp Phe Glu Glu  Asp Phe Asp
    1595                 1600                 1605

Pro Ser  Val Lys Trp Lys Leu  Leu Pro Gln Glu Glu  Pro
    1610                 1615                 1620
```

<210> 3
<211> 14
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 3
cagaagtcaa atag 14

<210> 4
<211> 16
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 4
caaaactata tatatg 16

<210> 5
<211> 22
<212> DNA
<213> synthetic oligonucleotide

<400> 5
catggtagag accaggagaa aa 22

<210> 6
<211> 21
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 6
ctcagtcatg tgagctgttg c 21

<210> 7
<211> 15
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 7
atgaggaaga gaaag 15

<210> 8
<211> 19
<212> DNA
<213> Artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 8
ttaaggctct tcctggggg 19

<210> 9
<211> 13
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 9
atacaaaggg ttc 13

<210> 10
<211> 12
<212> DNA
<213> artificial sequence

<220>
<223> Synthetic oligonucleotide

<400> 10
ctgaggatct cc 12